日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: １９９９年１１月１９日

出 願 番 号
Application Number: 平成１１年特許願第３３０２３６号

出 願 人
Applicant (s): 株式会社東芝

CERTIFIED COPY OF
PRIORITY DOCUMENT

２０００年１０月２０日

特 許 庁 長 官
Commissioner,
Patent Office 及 川 耕 造

【書類名】　特許願

【整理番号】　A009906696

【提出日】　平成11年11月19日

【あて先】　特許庁長官　殿

【国際特許分類】　G06F 15/00

【発明の名称】　構造化文書検索方法、構造化文書検索装置及び構造化文書検索システム

【請求項の数】　14

【発明者】

【住所又は居所】　神奈川県川崎市幸区柳町７０番地　株式会社東芝柳町工場内

【氏名】　服部　雅一

【発明者】

【住所又は居所】　神奈川県川崎市幸区柳町７０番地　株式会社東芝柳町工場内

【氏名】　野々村　克彦

【発明者】

【住所又は居所】　神奈川県川崎市幸区柳町７０番地　株式会社東芝柳町工場内

【氏名】　金輪　拓也

【特許出願人】

【識別番号】　000003078

【氏名又は名称】　株式会社　東芝

【代理人】

【識別番号】　100058479

【弁理士】

【氏名又は名称】　鈴江　武彦

【電話番号】　03-3502-3181

【選任した代理人】

【識別番号】　100084618

【弁理士】

【氏名又は名称】　村松　貞男

【選任した代理人】

【識別番号】　100068814

【弁理士】

【氏名又は名称】　坪井　淳

【選任した代理人】

【識別番号】　100092196

【弁理士】

【氏名又は名称】　橋本　良郎

【選任した代理人】

【識別番号】　100091351

【弁理士】

【氏名又は名称】　河野　哲

【選任した代理人】

【識別番号】　100088683

【弁理士】

【氏名又は名称】　中村　誠

【選任した代理人】

【識別番号】　100070437

【弁理士】

【氏名又は名称】　河井　将次

【手数料の表示】

【予納台帳番号】　011567

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　構造化文書検索方法、構造化文書検索装置及び構造化文書検索システム

【特許請求の範囲】

【請求項１】

論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行う構造化文書検索方法であって、

前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成し、

前記構造化文書データベースにおける実データに関するインデックス情報を利用して、前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成し、

前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足する検索結果を求めることを特徴とする構造化文書検索方法。

【請求項２】

前記検索プランの生成においては、前記インデックス情報を利用しながら前記検索グラフを巡回することによって最適な検索プランを生成することを特徴とする請求項１に記載の構造化文書検索方法。

【請求項３】

前記検索グラフ中において評価可能な部分グラフを優先的に評価する戦略に基づいて前記検索グラフを巡回することを特徴とする請求項に２記載の構造化文書検索方法。

【請求項４】

前記検索プランの生成が全て完了した後に、該検索プランの実行を行うことを特徴とする請求項１に記載の構造化文書検索方法。

【請求項５】

前記検索プランの生成および実行を交互に繰り返し行うことを特徴とする請求項１に記載の構造化文書検索方法。

【請求項６】

前記構造化文書データベースは、要素名称および要素値に関係する階層構造を含み、

前記検索要求は、前記要素名称および前記要素値に関係する検索条件を含み、

前記インデックス情報は、前記構造化文書データベースにおける前記要素値の生起位置を特定する情報を含むデータ生起インデックスと前記構造化文書データベースにおける前記要素名称の生起位置を特定する情報を含む要素名称生起インデックスとの少なくとも一方を含むことを特徴とする請求項１に記載の構造化文書検索方法。

【請求項７】

前記要素名称生起インデックスは、前記要素名称の生起位置を、前記要素名称の発生する部分構造の一階層上位の親要素によって指し示した情報を含むこと特徴とする請求項６に記載の構造化文書検索方法。

【請求項８】

前記検索プランの生成においては、

ルール適用条件を示す情報と前記検索プランを構成すべき検索処理の内容を指示する情報とを含むプラン生成ルールが複数登録されたプラン生成ルールベースに基づき、プラン生成ルールを選択し、該プラン生成ルールを前記検索グラフの該当する要素に対して適用するとともに、該プラン生成ルールに含まれる検索処理を、前記検索プランを構成する１つの検索処理として決定し、

前記プラン生成ルールが適用された結果として影響が及ぶ前記検索グラフの要素に関して、プラン生成ルールの選択および適用ならびに前記検索プランにおいて後続させる検索処理の決定を行うことを、繰り返し行うことによって、

前記検索プランを生成していくことを特徴とする請求項２に記載の構造化文書検索方法。

【請求項９】

前記プラン生成ルールには、前記インデックス情報を加味して決定されるコスト情報が付与されており、

前記コスト情報を考慮して、動的に、適用すべきプラン生成ルールを選択することを特徴とする請求項８に記載の構造化文書検索方法。

【請求項１０】

前記プラン生成ルールベースにおける前記プラン生成ルールを任意に登録および削除可能としたことを特徴とする請求項８に記載の構造化文書検索方法。

【請求項１１】

前記検索グラフの生成においては、前記検索要求の記述を構文解析した結果に基づいて前記検索グラフの生成を行うことを特徴とする請求項１に記載の構造化文書検索方法。

【請求項１２】

論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行う構造化文書検索装置であって、

前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成する手段と、

前記構造化文書データベースにおける実データに関するインデックス情報を利用して、前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成する手段と、

前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足する検索結果を求める手段とを備えたことを特徴とする構造化文書検索装置。

【請求項１３】

論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行うためのプログラムであって、

前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成させ、

前記構造化文書データベースにおける実データに関するインデックス情報を利用して、前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成させ、

前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足する検索結果を求めるさせるためのプログラムを記録したコンピュータ読取り可能な記録媒体。

【請求項１４】

論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行う構造化文書検索システムであって、

前記構造化文書データベースの実データを記憶する手段と、

前記構造化文書データベースにおける実データに関するインデックス情報を記憶する手段と、

外部から前記検索要求を受け付ける手段と、

受け付けた前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成する手段と、

前記構造化文書データベースにおける実データに関するインデックス情報を利用して、前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成する手段と、

前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足する検索結果を求める手段と、

前記検索結果を外部へ出力する手段とを備えたことを特徴とする構造化文書検索システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行う構造化文書検索方法、構造化文書検索装置及び構造化文書検索システムに関する。

【０００２】

【従来の技術】

従来から文書データベースに対する検索要求を指定する方法としてキーワード指定がある。ユーザが検索要求をキーワード列という形式で文書データベースに要求すると、キーワード列を含んでいる文書群を返すというものである。

【０００３】

このような素朴で原始的な検索要求方式は、全文検索エンジンなどに広く適用されているが、それゆえ、（１）必要以上の文書群が検索されてしまうという低

い精度の問題や、（２）利用部分以外のデータまで含んだ文書がデータ単位であるという粒度の問題がある。

【０００４】

近年、ＳＧＭＬ（Ｓｔａｎｄａｒｄ Ｇｅｎｅｒａｌｉｚｅｄ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）やＸＭＬ（ｅＸｔｅｎｓｉｂｌｅ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）などの構造化文書のための構造化文書規約が提案され、文書構造に基づいた検索要求の指定により、（１）従来のキーワード検索よりも精度の高い検索と、（２）利用部分だけのデータが得られるという木目細かい検索が可能になっている。しかしながら、この場合、予め文書構造を固定的なものに統一する必要があり、後から文書構造の変更ができない、あるいはデータ毎に文書構造を変えることができないという欠点がある。

【０００５】

一方、ＲＤＢ（Ｒｅｌａｔｉｏｎａｌ ＤａｔａＢａｓｅ）では、表の構造に基づいた検索要求をＳＱＬ言語によって指定することができる。ＳＱＬは、ＡＮＳＩ Ｘ３，１、およびＩＳＯ／ＴＣ９７／ＳＣ２１／ＷＧ３ Ｎ１１７（１９８７）において標準化されたＲＤＢの問合せ言語である。しかしながら、文書構造はそのまま表形式に変換することは困難であり、ＲＤＢを文書データベースとしてそのまま用いることはできない。

【０００６】

さらに、ＳＧＭＬやＸＭＬなどの構造化文書データベースに対するＯＯＤＢ（Ｏｂｊｅｃｔ Ｏｒｉｅｎｔｅｄ ＤａｔａＢａｓｅ）で用いられた検索言語を適用する方法が考えられる。構造化文書は階層的な構造を持つため、各構成要素をオブジェクトとみなしたＯＯＤＢと親和性が高いと考えられる。しかしながら、ＯＯＤＢでは、文書構造はあらかじめスキーマにより決定されていなければならず、子要素の任意繰り返しなど、オブジェクトモデルでモデル化するのは困難であり、オブジェクト指向データベースを文書データベースとしてそのまま用いることはできない。

【０００７】

このような問題を解決するために、文書リポジトリに対して、ＳＱＬへ構造化

文書に適した拡張機能を追加した言語処理部を装備することが考えられている。構造化文書に適した拡張機能には、階層的な構造上の構成要素を特定するパス指定が第一に挙げられる。さらに、階層的な構造上の構成要素を特定するパスに正規表現などの曖昧性を含んだ曖昧パス指定や、階層的な構造のパターンを指定する構造パターン指定など、構造化文書が持つ構造的な揺らぎを吸収するような機能がＳＱＬをベースに拡張されている。

【０００８】

これらの特徴を持った検索要求を指定でき、かつ検索処理できる方式を提案しているものに、特開平６－２０３０７８号公報、特開平６－３０１７２１号公報、特開平１１－１５８４３号公報がある。

【０００９】

特開平６－２０３０７８号公報（情報検索方法およびその装置）では、階層構造を全展開したパス集合を文字列表としてＲＤＢに格納する方式を提案している。構造化文書を検索するとき、文字列表のパスを検索文の曖昧パスと文字列比較するＳＱＬを発行することで、階層的な構造上の構成要素を特定している。この方式の問題点は、登録された文書数が増大すると、階層構造を全展開した文字列表が膨大なサイズになってしまうことである。

【００１０】

特開平６－３０１７２１号公報（全文データベース検索方式）では、構成要素タイプをあらかじめ決めておき、その階層構造の親子関係や実データへのリンクなどを構成要素タイプ毎に構造情報としてＲＤＢ化する方法を提案している。構造化文書を検索するとき、検索要求をＳＱＬ文に変換している。この方式の問題点は、ルート要素から始まって、親要素から子供要素群へと展開して、階層的な構造上の構成要素を特定する検索処理方式であるため、登録された文書数が増大し階層木の深さと幅が増大すると、検索処理に要する計算量が膨大なものになってしまうことである。ＲＤＢの結合で展開処理を行っているため、実装システムは想像を超えた応答時間が予想される。特に曖昧パスが指定されたときは、その傾向が激しくなる。

【００１１】

特開平１１－１５８４３号公報（ＳＧＭＬ文書検索装置およびＳＧＭＬ文書検索方法）でも、構成要素タイプをあらかじめ決めておき、構成要素タイプ毎にデータを文字列結合した文書テーブルを作成しておく。構造化文書を検索するとき、検索要求をＳＱＬ文に変換している。この方式の問題点は、構成要素タイプ毎にデータをただ単に文字列結合するため、１段レベルのパスしか指定できないことである。さらに、あらかじめ文書構造が決まっていなければならず、文章が持つ階層構造に対する柔軟な検索要求は発行できない、などの問題点も抱えている。

【００１２】

これらの方式では、データに対するインデックスと構造に関するインデックスを適切に組み合わせて、検索処理に要する計算量を抑えるようになっておらず、ＲＤＢのような最適化を入れにくい仕組みとなっている。

【００１３】

【発明が解決しようとする課題】

以上説明したように従来の技術では、（１）（曖昧パスを含む）文書が持つ階層構造に対する多様な検索指定を行うことと、（２）検索処理に要する計算量を膨大なものとしないことというトレードオフの関係にある２つの要求を同時に満足させることは困難であった。

【００１４】

本発明は、上記事情を考慮してなされたもので、検索処理に要する計算量の増大を伴わずに、（曖昧パスを含む）文書が持つ階層構造に対する多様な検索指定を行うことを可能とした、構造化文書検索方法、構造化文書検索装置及び構造化文書検索システムを提供することを目的とする。

【００１５】

【課題を解決するための手段】

本発明（請求項１）は、論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行う構造化文書検索方法であって、前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成し、前記構造化文書データベースにおける実データに関するインデックス情報を利用して、

前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成し、前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足する検索結果を求めることを特徴とする。

【００１６】

好ましくは、前記検索プランの生成においては、前記インデックス情報を利用しながら前記検索グラフを巡回することによって最適な検索プランを生成するようにしてもよい。

【００１７】

好ましくは、前記検索グラフ中において評価可能な部分グラフを優先的に評価する戦略に基づいて前記検索グラフを巡回するようにしてもよい。

【００１８】

好ましくは、前記検索プランの生成が全て完了した後に、該検索プランの実行を行うようにしてもよい。

【００１９】

好ましくは、前記検索プランの生成および実行を交互に繰り返し行うようにしてもよい。

【００２０】

好ましくは、前記構造化文書データベースは、要素名称および要素値に関係する階層構造を含み、前記検索要求は、前記要素名称および前記要素値に関係する検索条件を含み、前記インデックス情報は、前記構造化文書データベースにおける前記要素値の生起位置を特定する情報を含むデータ生起インデックスと前記構造化文書データベースにおける前記要素名称の生起位置を特定する情報を含む要素名称生起インデックスとの少なくとも一方を含むようにしてもよい。

【００２１】

好ましくは、前記要素名称生起インデックスは、前記要素名称の生起位置を、前記要素名称の発生する部分構造の一階層上位の親要素によって指し示した情報を含むようにしてもよい。

【００２２】

好ましくは、前記検索プランの生成においては、ルール適用条件を示す情報と前記検索プランを構成すべき検索処理の内容を指示する情報とを含むプラン生成ルールが複数登録されたプラン生成ルールベースに基づき、プラン生成ルールを選択し、該プラン生成ルールを前記検索グラフの該当する要素に対して適用するとともに、該プラン生成ルールに含まれる検索処理を、前記検索プランを構成する１つの検索処理として決定し、前記プラン生成ルールが適用された結果として影響が及ぶ前記検索グラフの要素に関して、プラン生成ルールの選択および適用ならびに前記検索プランにおいて後続させる検索処理の決定を行うことを、繰り返し（伝播的に）行うことによって、前記検索プランを生成していくようにしてもよい。

【００２３】

好ましくは、前記プラン生成ルールには、前記インデックス情報を加味して決定されるコスト情報が付与されており、前記コスト情報を考慮して、動的に、適用すべきプラン生成ルールを選択するようにしてもよい。

【００２４】

好ましくは、前記プラン生成ルールベースにおける前記プラン生成ルールを任意に登録および削除可能とするようにしてもよい。これによって、検索プランの生成をカスタマイズすることができる。

【００２５】

好ましくは、前記検索グラフの生成においては、前記検索要求の記述を構文解析した結果に基づいて前記検索グラフの生成を行うようにしてもよい。

【００２６】

また、本発明（請求項１２）は、論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行う構造化文書検索装置であって、前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成する手段と、前記構造化文書データベースにおける実データに関するインデックス情報を利用して、前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成する手段と、前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足す

る検索結果を求める手段とを備えたことを特徴とする。

【００２７】

また、本発明（請求項１３）は、論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行うためのプログラムであって、前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成させ、前記構造化文書データベースにおける実データに関するインデックス情報を利用して、前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成させ、前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足する検索結果を求めるさせるためのプログラムを記録したコンピュータ読取り可能な記録媒体である。

【００２８】

また、本発明（請求項１４）は、論理構造を持つ構造化文書データベースに対して、文書の論理構造を含む検索要求に基づいて検索を行う構造化文書検索システムであって、前記構造化文書データベースの実データを記憶する手段と、前記構造化文書データベースにおける実データに関するインデックス情報を記憶する手段と、外部から前記検索要求を受け付ける手段と、受け付けた前記検索要求に基づいて、文書の構造情報を含む検索グラフを生成する手段と、前記構造化文書データベースにおける実データに関するインデックス情報を利用して、前記検索グラフから、前記構造化文書データベースに対する検索処理手順を示す検索プランを生成する手段と、前記構造化文書データベースを検索対象として前記検索プランを実行することによって、前記検索要求を満足する検索結果を求める手段と、前記検索結果を外部へ出力する手段とを備えたことを特徴とする。

【００２９】

なお、方法に係る本発明は装置／システムに係る発明としても成立し、装置／システムに係る本発明は方法に係る発明としても成立する。

【００３０】

また、方法または装置／システムに係る本発明は、コンピュータに当該発明に相当する手順を実行させるための（あるいはコンピュータを当該発明に相当する

手段として機能させるための、あるいはコンピュータに当該発明に相当する機能を実現させるための）プログラムを記録したコンピュータ読取り可能な記録媒体としても成立する。

【００３１】

本発明では、必要に応じて（要素名称生起やデータ生起など）様々なインデックス情報や文書の階層構造に関する情報を有効に利用して、（文書の構造情報を含んだ）検索グラフを最適に巡回することで最適な検索プランを生成し、これを実行する。すなわち、本発明によれば、存在するインデックス情報を動的に用いながら、最適な検索プラン生成・実行することが可能になる。また、本発明によれば、（構造照合やデータ比較を組み合わせた）多様な検索指定が行われていても、最適な検索プランを生成することで、検索処理に要する計算量を抑えることができる。

【００３２】

このように本発明によれば、（１）（曖昧パスを含む）文書が持つ階層構造に対する多様な検索指定を行いながら、（２）検索処理に要する計算量を膨大なものとしない、という両要求を同時に満足し、論理構造を持った構造化文書データベースに対して文書の論理構造を含めた検索要求文で検索するサービスを実現することができる。

【００３３】

【発明の実施の形態】

以下、図面を参照しながら発明の実施の形態を説明する。

【００３４】

本発明を適用可能な構造化文章として、例えば、ＳＧＭＬ（Ｓｔａｎｄａｒｄ Ｇｅｎｅｒａｌｉｚｅｄ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）やＸＭＬ（ｅＸｔｅｎｓｉｂｌｅ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）で記述された文書が挙げられる。ＳＧＭＬとは、ＩＳＯ（国際標準機構）で定められた規格である。ＸＭＬとは、Ｗ３Ｃ（ワールドワイドウェブコンソーシアム）にて定められた規格である。それぞれ文書を構造化することを可能とする構造化文書規約である。

【００３５】

ＳＧＭＬやＸＭＬを用いた文書の構造の表現にはタグが用いられる。タグには、開始タグと終了タグがあり、文書構造情報の構成要素を開始タグと終了タグで囲むことにより、文書中の文章の区切りと、その文書が構造上どの構成要素に属するのかとを明確にする。ここで、開始タグは「要素名称」を記号「＜」と「＞」で閉じたものであり、終了タグは「要素名称」を記号記号「＜」と「／＞」で閉じたものである。タグに続く構成要素の内容が、テキストまたは子供の構成要素の繰り返しである。また、開始タグには「＜要素名称　属性＝“属性値”＞」のように属性情報を設定することができる。

【００３６】

以下では具体例としてＸＭＬを用いて説明するものとする。

【００３７】

また、データベースの内容の具体例として特許出願に関する情報を用い、検索の具体例として特許出願に関する情報の検索を用いるものとする。なお、具体例を用いた説明で「特許」という場合は、「特許出願に関する（もの）」というような意味で用いているものとする。

【００３８】

図１に、本発明の一実施形態に係る構造化文書データベース・システムのシステム構成を示す。

【００３９】

本システムは、要求制御部１、格納処理部２、検索処理部３、データファイル４、インデックスファイル５を含む。

【００４０】

本システム構成は、ソフトウェアを用いて実現可能である。なお、データファイル４、インデックスファイル５は、例えば外部記憶装置を用いて構成される。

【００４１】

要求制御部１は、ユーザからの検索要求や格納要求など構造化文書データベースへの要求を処理し、検索処理部３や格納処理部２へ処理を渡す処理部である。検索要求と格納要求は要求受付部１１でメッセージとして受け取る。受け取ったメッセージについて要求処理部１２で検索要求か格納要求かの分別を行い、検索

処理部３による検索処理あるいは格納処理部２による格納処理を呼び出す。また、検索処理部３から渡された検索結果は、結果処理部１３にて整形されて、要求元のユーザに返される。

【００４２】

検索処理部３は、検索要求を解析し、検索要求を満足する検索結果を生成する処理部である。検索要求構文解析部３１にて、検索要求から字句切り出しや要求文の構造抽出を行い、検索グラフ生成部３２にて、検索グラフを生成する。検索プラン生成部３３にて、生成された検索グラフから検索プランを生成し、検索プラン実行部３４にて、生成された検索プランを実行し、検索要求を満足する検索結果を生成する。検索結果は、要求制御部１に渡される。

【００４３】

格納処理部２は、格納要求を解析し、構造化文書を格納する処理部である。格納要求構文解析部２１にて、構造化文書から字句切り出しや構造化文書の構造抽出を行う。データ格納部２２にて、構造化文書のデータや構造データをデータファイル４に格納し、インデックス格納部２３にて、構造化文書のデータや構造データに対するインデックスをインデックスファイル５に格納する。なお、インデックスファイルの作成・更新は、格納すべき構造化文書が入力されるごとに行ってもよいし、適宜まとめて行ってもよい（検索の効率化のためには、前者の方が好ましい）。

【００４４】

図２に、構造化文書の一例を示す。

【００４５】

図２は、構造化文書の一例として「特許」情報の例を示したものである（ＸＭＬで記述した例である）。

【００４６】

「特許」タグ（すなわち、＜特許＞と＜／特許＞の対；他も同様の意味である）で囲まれた内部には、「名称」タグで囲まれた「名称」情報、「出願人」タグで囲まれた「出願人」情報、「出願番号」タグで囲まれた「出願番号」情報、「出願日」タグで囲まれた「出願日」情報、「要約」タグで囲まれた「要約」情報

、「キーワード」タグで囲まれた「キーワード」情報が存在する。

【００４７】

「出願日」情報は、さらに、「年」タグ囲まれた「年」情報、「月」タグ囲まれた「月」情報、「日」タグで囲まれた「日」情報により構成される。なお、「出願日」情報は、「年号」情報をさらに含んでもよい。あるいは、「年」情報を西暦で表してもよい。

【００４８】

また、「キーワード」情報としては、１または複数個のものを指定することができる（図２の例では２個のキーワード「ＸＭＬ」、「検索」が指定されている）。

【００４９】

この「キーワード」情報のように、ＸＭＬなどの構造化文書では、任意の構成要素の繰り返しを含んでいたり、さらには文書構造があらかじめ決まっていない（ＲＤＢやＯＯＤＢのスキーマ定義では定義できない）のが通常である。

【００５０】

なお、「特許」情報には、「公開番号」情報や、「特許番号」情報、あるいはその他の種々の情報を含めることができる。

【００５１】

図３および図４に、本実施形態で必要に応じて検索で使用される概念階層を構造化文書で表現した例を示す。図３および図４の例は、「概念」情報をＸＭＬで記述したものである。

【００５２】

図３の「概念」情報の例は、いわゆる特許調査における特許文書の内容を分類するための一つの分類軸として用いる「情報モデル」を概念階層で表現している。「概念」タグで囲まれた「概念」情報は、入れ子構造を持った文書構造を持っている。つまり、図３の例では、概念「情報モデル」の子供概念として、概念「ドキュメント」、概念「リレーション」、概念「オブジェクト」が存在している。また、概念「ドキュメント」の子供概念として、概念「構造化ドキュメント」、概念「非構造化ドキュメント」が存在し、さらに、概念「構造化ドキュメント

」の子供概念として、概念「ＸＭＬ」、概念「ＳＧＭＬ」が存在している。

【００５３】

図４の概念階層の記述例は、図３とは異なる分類軸「情報操作」を概念階層で表現している。図４の例では、概念「情報操作」の子供概念として、概念「検索」、概念「格納」、概念「加工」、概念「流通」が存在している。

【００５４】

図５に、本実施形態における構造化文書データベースの概念的な構造の例を示す。

【００５５】

構造化文書を集めた構造化文書データベースは、例えばＵＮＩＸのディレクトリ構造のように階層的に格納されていることを指定している。

【００５６】

構造化文書データベースの階層木の各ノード（図５では番号が付され円形で示されたもの）を、ドキュメントノードと呼ぶ。なお、以下では、ドキュメントノードをＤノードと呼ぶ。

【００５７】

任意のＤノード以下の部分階層木は、構造化文書データベースから切り出された構造化文書を示している。

【００５８】

各Ｄノードには、オブジェクトＩＤ（図５では円内部に記述されたもの）が割り当てられる。オブジェクトＩＤは、構造化文書データベース内ではユニークな数値を持つものとする。

【００５９】

図５の例では、階層木のルートとなるドキュメントノード（根Ｄノード）に、それが根Ｄノードであることを特定可能なオブジェクトＩＤ「＃０」が割り当てられるものとしている。

【００６０】

図５の例において、根Ｄノードすなわち「＃０」のＤノードからは、「ｒｏｏｔ」タグを先頭に持つ「＃１７」のＤノードへリンクが張られている。「＃１７

」Ｄノードからは、「ＩＲ特許」タグを先頭に持つ「＃２１」Ｄノード、「ＤＢ特許」タグを先頭に持つ「＃４５」Ｄノード、「概念」タグを先頭に持つ「＃７８」Ｄノードへのリンクがそれぞれ張られている。なお、ＩＲ特許とは、例えば、ＩＲ技術に関係する発明をその明細書中に含む特許出願というような意味である（ＤＢ特許、ＯＯＤＢ特許、ＲＤＢ特許についても同様である）。

【００６１】

図２に例示された「特許」情報は、「＃９０２」のＤノード以下の部分階層木に対応しており、「名称」タグあるいは「キーワードタグ」などを先頭に持つ各々の末端のＤノード（＃９０３～＃９０５、＃９０７～＃９１２）からは、「情報検索装置」、「Ｔ社」、「特願平１０－××××××」、「１０」、「３」、「１２」、「情報の提示形式の変更が～（以下、省略）」、「ＸＭＬ」、「検索」などの文字列（要素値）へのリンクがそれぞれ張られている。

【００６２】

ところで、「＃６３９」のＤノード以下の部分階層木も一つの「特許」情報に対応する部分であるが、根Ｄノードからみて「＃９０２」Ｄノードと「＃６３９」Ｄノードとは階層の深さが異なっている。このように、根Ｄノードから「特許」情報に対応するＤノードまでの階層関係は任意に設定することが可能である。

【００６３】

すなわち、図５に示されているように「特許」情報は、「＃９０２」Ｄノードや「＃６３９」Ｄノードなどのように階層木上の任意の部分に発生し得る。これが構造化文書データベースの特徴である。そのため、階層木上の任意の部分に発生した「特許」情報を検索したいという検索要求がある。

【００６４】

なお、本実施形態では、図５に示すように、図３や図４のような「概念」情報も構造化文書データベース内に併せて保持することができる（例えば、「＃７８」Ｄノード下位の以下の部分階層木に含まれる）。

【００６５】

図６に、本実施形態における構造化文書データベースへの構造化文書の格納コマンドの一例を示す。

【００６６】

コマンド名「Ｉｎｓｅｒｔ」の後に、格納先「“ｒｏｏｔ／ＩＲ特許”」、格納データ「“＜特許＞～（中略）～＜／特許＞”」の２つのパラメータが存在する。この記述は、格納先として、「ｒｏｏｔ」タグを先頭に持つ部分階層木から辿って、「ＩＲ特許」タグを先頭に持つ部分階層木の先頭要素に、格納データ「“＜特許＞～（中略）～＜／特許＞”」を挿入することを意味する。「“ｒｏｏｔ／ＩＲ特許”」を文書パスと呼ぶ。

【００６７】

図６に例示した格納コマンドを実行した結果として、図５に例示した概念的な構造の「＃９０２」のＤノード以下の部分階層木が作られることになる。

【００６８】

「Ｉｎｓｅｒｔ」コマンド名を持つ格納要求は、図１の要求制御部１にて受理され、格納処理部２による構文解析（２１）を経て、データ格納（２２）とインデックス格納（２３）が行われる。

【００６９】

図７に、構造化文書データベースへの検索コマンドの一例を示す。

【００７０】

図７の例は、検索コマンドをＳＱＬに似たＳｅｌｅｃｔ文で表現したもので、『構造化文書データベース中に出現する「特許」情報のうちその「キーワード」情報として「検索」を持つものについて、「出願番号」情報を抽出し、それを「文献」情報として出力せよ』という検索要求を意味している。

【００７１】

「Ｗｈｅｒｅ」句が条件部分を示しており、「Ｆｒｏｍ」句が文書パス指定部分を示しており、「Ｓｅｌｅｃｔ」句が情報抽出部分を示している。「＄１」、「＄２」はデータが束縛される変数である。

【００７２】

「＜＊／特許＞」のように要素名称の前が「＊」で始まっていれば、指定された文書パスの任意子孫の「特許」にマッチすることができる。「ｒｏｏｔ／＊／特許」のように曖昧な文書パスが、『階層木上の任意の部分に発生した「特許」

情報を検索したいという検索要求』に対応する。

【００７３】

例えば、図５において、「＃９０２」Ｄノード以下の部分階層木に対応する「特許」情報について、「＃９１２」Ｄノードからリンクされた「検索」が条件を満たし、「＃９０５」Ｄノードからリンクされた「特願平１０－××××××」が検索結果となる。

【００７４】

図８に、構造化文書データベースへの検索コマンドの他の例を示す。

【００７５】

この例は、『構造化文書データベース中に出現する「特許」情報のうち、その「キーワード」情報として、概念「ドキュメント」に属する内容（図３では、概念の名前の属性値（文字列）に一致する要素値を持つ「特許」情報について、「出願番号」情報を抽出し、それを「文献」情報として出力せよ』という検索要求を意味している。

【００７６】

この例では、「特許」情報と「概念」情報の２つを参照し、それぞれ「キーワード」と「名前」に対して、同一変数「＄ｘ２」が割り当てられている。これは２つの情報の結合処理を意味している。

【００７７】

例えば、図５において、「概念」情報が図３のようであるとすると、「＃９１１」のＤノードからリンクされた「ＸＭＬ」が図３のように概念「ドキュメント」に属するので、「＃９０５」のＤノードからリンクされた「特願平１０－××××××」が検索結果となる。

【００７８】

図９に、構造化文書データベースへの検索コマンドのさらに他の例を示す。

【００７９】

この例は、『構造化文書データベース中に出現する「特許」情報に対して、概念「情報モデル」での分類と概念「情報操作」での分類の２分類軸を設定して、「出願番号」と「情報モデル」軸と「情報操作」軸とを抽出して「文献」情報と

して検索せよ』を意味している。

【００８０】

この例では、「特許」情報と「概念」情報の２つを参照し、それぞれ「キーワード」と「名前」に対して、同一変数「＄ｘ２」が割り当てられている。これも２つの情報の結合処理を意味している。

【００８１】

「情報モデル」軸を取り出す部分では、「特許」情報の「キーワード」情報「＄ｘ２」が文書パス「ｒｏｏｔ」以下の概念「情報モデル」の任意子孫の「概念」情報にマッチするものを探索し、概念「情報モデル」の１つ子供の概念に置き換えて「＄ｘ３」とする処理が組み込まれている。「情報操作」軸を取り出す部分も同様に、「特許」情報の「キーワード」情報「＄ｘ２」が文書パス「ｒｏｏｔ」以下の概念「情報操作」の任意子孫の「概念」情報にマッチするものを探索し、概念「情報操作」の１つ子供の概念に置き換えて「＄ｘ４」とする処理が組み込まれている。

【００８２】

例えば、図５において、「概念」情報が図３および図４のようであるとすると、「＃９１１」のＤノードからリンクされた「ＸＭＬ」が図３のように概念「情報モデル」に属し、かつ、「＃９１２」のＤノードからリンクされた「検索」が図４のように概念「情報操作」に属するので、「＃９０５」のＤノードからリンクされた「特願平１０－××××××」と、図３の概念「情報モデル」の１つ子供の概念「ドキュメント」と、図４のように概念「情報操作」の１つ子供の概念「検索」とが検索結果となる。

【００８３】

図１０に、図９の検索要求を処理した検索結果の一例を示す。図１０に例示されるように、検索結果もＸＭＬで表現することができる。

【００８４】

図９で示された検索要求は、図１の要求制御部１にて受理され、検索処理部３にて構文解析（３１）、検索グラフ生成（３２）、検索プラン生成（３３）、検索プラン実行（３４）などの一連の処理を経て、要求制御部１の結果処理部１３

にて整形されて、図１０に示すような検索結果が得られる。

【００８５】

先にも述べたように、「特許」情報に対して、概念「情報モデル」での分類と概念「情報操作」での分類の２分類軸を設定して、「出願番号」情報とともにまとめられて「文献」情報のリストとして表示されている。例えば、第一の「文献」情報では、『「特願平１０－××××××」の特許が「ドキュメント」×「検索」で分類されている』ことを意味している。

【００８６】

以下では、検索処理部３における処理についてより詳しく説明する。

【００８７】

図１１および図１２に、図９の検索要求に対して検索グラフ生成部３２が生成する検索グラフの一例を示す（なお、図１１、図１２は便宜上、同一の検索グラフの一部を省略したものであって、すなわち、図１１は同一の検索グラフのＣＯＮ以下の部分を省略したものであり、図１２は同一の検索グラフのＡＮＤ以下の部分を省略したものである）。

【００８８】

図１１および図１２に示されるように、検索グラフは、双方向リンク（図中の両方向の矢印）とノード（図中の円形、四角形、六角形）を含むネットワークを形成する。

【００８９】

図１１および図１２において、四角形で示されるノードは、具体的なデータ（文字列）を表している。四角形で示されるノードを除く各ノードを、検索グラフノード（以下、Ｇノード）と呼ぶ。すなわち、Ｇノードは、六角形で示されるＧノードと円形で示されるＧノードの２種類から構成される。

【００９０】

円形で示されるＧノードは、変数を表すＧノードであり、「＄＿」で始まる文字列を持っている。変数を表すＧノードは、内部的に生成された変数と、それ以外の「＄ｘ１」など検索要求のＳｅｌｅｃｔ文に含まれている変数とに分類できる。

【００９１】

六角形で示されるＧノードは、「ＱＵＥＲＹ」のＧノード、「ＡＮＤ」のＧノード、「ＴＡＧ」のＧノード、「ＡＴＴ」のＧノード、「ＶＡＬ」のＧノードなどから構成されている。

【００９２】

「Ｑｕｅｒｙ」Ｇノードは図９のような検索要求のＳｅｌｅｃｔ文全体に対応しており、「ＡＮＤ」ＧノードはＷｈｅｒｅ句に対応しており、「ＣＯＮ」ＧノードはＳｅｌｅｃｔ句に対応している。Ｗｈｅｒｅ句以下の複合的な条件部分は、「ＡＮＤ」Ｇノードと「ＡＮＤ」Ｇノードから出ているサブネット群が対応している。

【００９３】

一例として、『指定された文書パス「ｒｏｏｔ」以下の任意の「特許」情報』の条件は、３つの「ＴＡＧ」Ｇノード（「ｒｏｏｔ」、「＊」、「特許」を持つＧノード列）で表現されている。２つの「ＴＡＧ」Ｇノードを繋ぐ変数Ｇノード（例えば、「＄＿１」や「＄＿２」）は、図５で示されるＤノードで束縛可能な変数である。例えば、「＄＿２」変数のＧノードは、右側２つの「ＴＡＧ」Ｇノードから解釈すると『指定された文書パス「ｒｏｏｔ」以下の任意の文書」を表し、左の「ＴＡＧ」Ｇノードと接続しているため、それも併せて解釈すると『指定された文書パス「ｒｏｏｔ」以下で「特許」タグを先頭に持つ任意の文書』を表す。

【００９４】

「ＡＴＴ」Ｇノード、「ＶＡＬ」Ｇノードは、それぞれ属性、要素データの関係を示している。

【００９５】

また、六角形で示されるＧノードには複数のリンクが接続する。

【００９６】

「ＱＵＥＲＹ」Ｇノードには、「ＡＮＤ」に接続するｏｐ１リンク、「ＣＯＮ」に接続するｏｐ２リンクがある。

【００９７】

「ＡＮＤ」Ｇノードには、「ＱＵＥＲＹ」に接続するｏｐ１リンク、「ＴＡＧ」群に接続するｏｐ２リンク群がある。

【００９８】

「ＴＡＧ」Ｇノードには、上位Ｇノードに接続するｏｐ１リンク（左側）、データに接続するｏｐ２リンク（下側）、下位Ｇノードに接続するｏｐ３リンク（右側）がある。

【００９９】

「ＡＴＴ」Ｇノードには、上位Ｇノードに接続するｏｐ１リンク（上側）、データに接続するｏｐ２リンク（右側）、下位Ｇノードに接続するｏｐ３リンク（下側）がある。

【０１００】

「ＣＯＮ」Ｇノードには、「ＱＵＥＲＹ」に接続するｏｐ１リンク、「ＴＡＧ」に接続するｏｐ２リンクがある。

【０１０１】

「ＶＡＬ」Ｇノードには、上位Ｇノードに接続するｏｐ１リンク（左側）、下位Ｇノードに接続するｏｐ２リンク（下側）がある。

【０１０２】

また、変数を表すＧノード（円形で示されるＧノード）には、他のＧノード群に接続するｏｐリング群がある。

【０１０３】

図９の例では、前述したように、「特許」情報と「概念」情報の２つを参照し、それぞれ「キーワード」と「名前」に対して、同一変数「＄ｘ２」が割り当てられている。「＄ｘ２」の変数Ｇノードは、３つの「ＶＡＬ」Ｇノードへのｏｐリンクとして接続し、逆に３つの「ＶＡＬ」Ｇノードからｏｐ２リンクとして接続されている。

【０１０４】

また、Ｓｅｌｅｃｔ句に対応している「ＣＯＮ」Ｇノードより下位のＧノードがネットワークを形成している。「特許」情報は、「出願番号」情報、「軸」属性が「情報モデル」の「分類」情報、「軸」属性が「情報操作」の「分類」情報

から構成されている。「＄ｘ１」、「＄ｘ３」、「＄ｘ４」などの変数Ｇノードは「ＡＮＤ」ＧノードであるＷｈｅｒｅ句が処理された後に変数値が確定し、束縛されて、図１０に示す結果となる。

【０１０５】

なお、図１１および図１２に例示された検索グラフは、例えばＹａｃｃ（Ｙｅｔ　Ａｎｏｔｈｅｒ　Ｃｏｍｐｉｌｅｒ－Ｃｏｍｐｉｌｅｒ）／Ｌｅｘ（ａ　ＬＥＸｉｃａｌ　ａｎａｌｙｚｅｒ　ｇｅｎｅｒａｔｏｒ）などの既存の構文解析プログラムジェネレータに、図９のような検索要求の記述を入力することによって生成することができる。

【０１０６】

次に、本実施形態で用いるインデックスファイルについて説明する。

【０１０７】

図１３に、インデックスファイルの一種である要素名称生起インデックスの概念的な構造の一例を示す。

【０１０８】

要素名称生起インデックスとは、構造化文書データベースに格納されている要素名称のリストと、各要素名称が先頭に発生する構造化文書の位置とを関連付けてインデックスファイル化したものである。

【０１０９】

例えば図５の構造化文書データベースのように、（「特許」情報に対応する）「特許」という要素名称が、Ｄノード群「＃９０２」、「＃６３９」、…により示される構造化文書において発生している場合、これをインデックス化すると、図１３に示すように、Ｄノード群「＃９０２」、「＃６３９」、…の親Ｄノード「＃２１」、「＃６７」、…が要素名称生起インデックスファイルに「特許」キーからのチェーンで格納される。

【０１１０】

このように親Ｄノードでインデックス化すると、インデックスファイルサイズを圧縮することができる。すなわち、親Ｄノードでインデックスすれば、子Ｄノードが増大しようとも、親Ｄノードで代用しているのでチェーンサイズは増大し

ない。これに対して、実Ｄノードをインデックス化すれば「特許」文書の格納数の増大とともにチェーンサイズはそれに比例して増加してしまう

図１４に、インデックスファイルの一種であるデータ生起インデックスの概念的な構造の一例を示す。

【０１１１】

データ生起インデックスとは、構造化文書データベースに格納されている文字列データのリストと各文字列データが発生する構造化文書の位置とを関連付けてインデックスファイル化したものである。

【０１１２】

例えば図５の構造化文書データベースのように、「検索」という文字列データが、Ｄノード群「＃９１２」、「＃６４７」、「＃６５０」、…により示される構造化文書にて発生している場合（なお、「＃６４７」のＤノードからリンクされるデータ中に検索という文字列が含まれているものとする）、これをインデックス化すると、図１４に示すように、Ｄノード群「＃９１２」、「＃６４７」、「＃６５０」…がデータ生起インデックスファイルに「検索」キーからのチェーンで格納される。

【０１１３】

なお、逆階層インデックスなど、その他のインデックスファイルを用いてもよい。逆階層インデックスとは、あるノードとその親ノードとの対応を格納したものである（あるノードからその親ノードを求めることができる）。

【０１１４】

次に、本実施形態の構造化文書データベースの検索プラン生成部３３について説明する。

【０１１５】

図１５に、検索プラン生成部３３の構成例を示す。図１５は、検索グラフ生成部３２にて生成された検索グラフを入力として実行プランリストを出力する検索プラン生成部３３の構成を表している。

【０１１６】

図中６は、後述するプラン生成ルールを格納したプラン生成ルール格納部であ

る（なお、プラン生成ルール格納部６は例えば外部記憶装置を用いて構成される）。

【０１１７】

候補Ｇノード登録部３３１は、図１１および図１２に例示したような検索グラフを構成する各Ｇノードを、候補Ｇノードリストへ登録する。

【０１１８】

Ｇノードルール発火チェック部３３２は、候補Ｇノードリストを構成するＧノードに対してプラン生成ルールの適用をチェックする。

【０１１９】

プラン生成ルール適用部３３３は、プラン生成ルールの適用可能な各Ｇノードについて、コスト最小のＧノードとプラン生成ルールとのペアを取り出し、プラン生成ルールを実行する。プラン生成ルールを実行した結果である実行プランが、実行プランリスト３３５へ追加される。

【０１２０】

また、プラン生成ルールによって値が具体化される可能性のある変数Ｇノード群について、候補Ｇノードリスト３３４に登録する。候補Ｇノードリスト３３４が空になるまで、これを繰り返す。

【０１２１】

このように、プラン生成ルールベースを使って、検索グラフの各要素に対してプラン生成ルールを適用し、適用された結果として影響のある検索グラフの各要素に対して再度プラン生成ルールを適用すること、すなわち、検索グラフを伝播的に巡回することで、格段に効率的な検索プランを実現することができる。

【０１２２】

なお、全ての検索プランの生成が完了した後に、生成された検索プランを実行するようにしてもよい。また、１つの検索プランの生成とその検索プランの実行とを一纏まりとして続けて行い、これを繰り返し実行する（すなわち、検索プランの生成とその実行を交互に繰り返し行う）ようにしてもよい。

【０１２３】

図１６に、プラン生成部３３で利用するプラン生成ルールの一例を示す。

【０１２４】

図１６のプラン生成ルール例は、１１個のルールをテーブル形式で記述したものである。

【０１２５】

各ルールには、ルール番号、適用可能なＧノードのクラス、適用コスト、適用条件（ＩＦ）部、アクション（ＴＨＥＮ）部の属性が存在する。

【０１２６】

コストは、０以上１以下のｆｌｏａｔ値を持ち、大きい数値を持つほど計算コストが大きくなることを意味するものとする。

【０１２７】

適用条件（ＩＦ）部においては、ＯＰ１～３は前述したリンクを表す。また、図１６中の「具」はそのリンクの変数Ｇノードが具体化されていることを表し、「未」は具体化されていないことを表し、「＊」はそのリンクの先のデータが「＊」であることを表し、「ＡＮＤ」はそのリンク先が「ＡＮＤ」Ｇノードであることを表す。なお、変数Ｇノードが具体化されているとは、『変数Ｇノードが取りうる値が枚挙可能な状態である』ことと定義する。

【０１２８】

適用条件（ＩＦ）部の「その他」の部分は、その他の適用条件を示す。例えば、ルール番号「０３」における「ＯＰ２に要素名称生起インデックスが存在」は、そのＯＰ２の具体化されている変数値と一致する要素名称が要素名称生起インデックスに存在することを適用条件とするものである。

【０１２９】

アクション（ＴＨＥＮ）部のオペレータは、詳しくは後述するように検索プラン実行部３４で実行されるアクションを示す。

【０１３０】

図１６において、例えば、ルール番号「０１」は、「ＴＡＧ」Ｇノードに対して適用可能で、コストが１．０であることを示している。さらに、適用条件が『ｏｐ１リンクの変数Ｇノードが「ＡＮＤ」であり、ｏｐ２リンクの変数Ｇノード（含むデータ）が具体化されていて、ｏｐ３リンクの変数Ｇノードが具体化され

ていない』ことである。またアクションが、『実行プランＰＡＴＨＩＮＳＴを生成する』ことを示している。

【０１３１】

また、例えば、ルール番号「０２」は、「ＴＡＧ」Ｇノードに対して適用可能で、コストが０．５であることを示している。さらに、適用条件が『ｏｐ１リンクの変数Ｇノードが具体化されていて、ｏｐ２リンクの変数Ｇノード（含むデータ）が具体化されていて、ｏｐ３リンクの変数Ｇノードが具体化されていない』ことである。またアクションが、『実行プランＰＡＴＨＥＸＰＡＮＤ１を生成する』ことを示している。

【０１３２】

図１７に、検索プラン実行部３４で利用されるオペレータの一例を示す。

【０１３３】

検索プラン実行部３４では、入力された実行プランリストを１つずつ取り出す（フェッチする）処理３４１と、実行する処理３４２とを繰り返し、その結果を検索結果として出力とする。

【０１３４】

各オペレータの処理内容は次の通りである。

【０１３５】

①ＰＡＴＨＩＮＳＴ　：　文書パス「ｒｏｏｔ」を取り出す処理。

②ＰＡＴＨＥＸＰＡＮＤ１　：　指定された要素名称をキーにして、上位Ｄノード群からキーにマッチするＤノード群を算出する処理。

③ＰＡＴＨＥＸＰＡＮＤ２　：　インデックス化されている要素名称をキーにして、構造化文書データベース内で発生する親子Ｄノード群を算出する処理。

④ＰＡＴＨＥＸＰＡＮＤ３　：　インデックス化されている要素名称をキーにして、子供Ｄノードから親Ｄノードを算出する処理。

⑤ＰＡＴＨＣＨＥＣＫ　：　２つのＤノード集合が与えられたとき、それらが指定された要素名称で親子関係にある２つのＤノードの組み合わせを算出する処理。

⑥ＪＯＩＮ　：　変数Ｇノードｘがｏｐリンクで接続している複数のＧノード

から具体化が進行して、ｘで重なり合ったときに行われる結合演算処理。

⑦ＶＡＬＵＥ　：　変数Ｇノードｘの要素データの候補を算出する処理。

⑧ＳＥＬＥＣＴ　：　変数Ｇノードｘに対する要素データを選択するときの比較演算処理。

⑨ＦＩＮＤ　：　インデックス化されている要素データの候補を算出する処理。

【０１３６】

図１８に、検索プラン生成部３３の処理手順の一例を示す。

【０１３７】

まず、候補Ｇノードリスト３３４と実行プランリスト３３５を空リストとして初期化する（ステップＳ１）。

【０１３８】

検索グラフを構成するＧノード全部を候補Ｇノードリスト３３４に登録する（ステップＳ２）。

【０１３９】

中間変数ｒｓｅｔを空リストとして初期化する（ステップＳ３）。

【０１４０】

候補Ｇノードリスト３３４が空リストであれば（ステップＳ４）、検索プラン生成部３３を終了する（ステップＳ４１）。

【０１４１】

候補Ｇノードリスト３３４が空リストでなければ（ステップＳ４）、空リストでない候補Ｇノードリスト３３４の各々の構成要素ｘに対して、ステップ５１からステップ５４まで繰り返す（ステップＳ５）。

【０１４２】

構成要素ｘに適用可能なプラン生成ルール群を検索する（ステップＳ５１）。

【０１４３】

検索されたプラン生成ルール群からＩＦ部を満足するプラン生成ルール群を選択する（ステップＳ５２）。

【０１４４】

選択されたプラン生成ルール群がなければ（ステップＳ５３）、候補Ｇノードリスト３３４から構成要素ｘを削除する（ステップＳ５３１）。

【０１４５】

選択されたプラン生成ルール群があれば（ステップＳ５３）、各々のプラン生成ルールｒに対してステップＳ５４１を適用する（ステップＳ５４）。

【０１４６】

プラン生成ルールｒのコストｃを計算し、ｒｓｅｔに＜ｘ，ｒ，ｃ＞を追加する（ステップＳ５４１）。

【０１４７】

続いて、ｒｓｅｔの各要素＜ｘ，ｒ，ｃ＞に対して、最小のコストｃを持つ要素＜ｘ１，ｒ１，ｃ１＞を選択する（ステップＳ６）。ここで、実行プランリスト３３５に、所定の事項を登録する。

【０１４８】

候補Ｇノードリスト３３４から構成要素ｘ１を削除する（ステップＳ７）。

【０１４９】

構成要素ｘ１に対してプラン生成ルールｒ１を実行し、更新可能性のあるＧノード（ｏｐ１，ｏｐ２，．．．などで繋がっている）を候補Ｇノードリスト３３４へ追加し、ステップＳ３に戻る（ステップＳ８）。

【０１５０】

以下では、構造化文書データベースへの検索コマンドの具体例を用いて検索グラフの生成から検索プランの生成、実行にわたってより具体的に説明する。

【０１５１】

図１９に、以下で用いる検索コマンド例を示す。この例は、『構造化文書データベース中に出現する「特許」情報に対して、下位の「名称」情報が「検索」という文字列を含んでいるならば、「名称」情報を抽出して「文献」情報として検索せよ』を意味している。

【０１５２】

図２０に、図１９の検索要求に対して検索グラフ生成部３３が生成する検索グラフの一例を示す。

【０１５３】

「＄１」変数Ｇノードは、『指定された文書パス「ｒｏｏｔ」以下で「＄２」変数Ｇノードよりも階層で上位に存在するＤノードに対するＧノード変数』を示している。

【０１５４】

「＄２」変数Ｇノードは、『「＄１」変数Ｇノードよりも階層で下位に存在し、「特許」要素名称で始まっているＤノードに対するＧノード変数』を示している。

【０１５５】

「＄３」変数Ｇノードは、『「＄２」変数Ｇノードから見て「特許」要素名称で始まるＤノードの子で『名称』要素名称で始まっているＤノードに対するＧノード変数』を示している。

【０１５６】

「＄４」変数Ｇノードは、『「＄３」変数Ｇノードから見て「名称」要素名称で始まるＤノードの子で要素データを指すＤノードに対するＧノード変数「＄ｘ１」を持つＤノードに対するＧノード変数』を示している。

【０１５７】

「＄ｘ１」変数Ｇノードは、『「＄４」変数Ｇノードから見て要素データを指すＤノードで「検索」という文字列を含むＤノードに対するＧノード変数』を示している。

【０１５８】

このようにＧノード同士は２項以上の多項間の制約関係を持っており、それらの変数Ｄノード群の取りうる値の組み合わせを制約充足的に解くことになる。

【０１５９】

図２１に、本実施形態の検索プラン生成部３３により生成される検索プランの一例を示す。

【０１６０】

図２１の検索プランは、図１９の検索要求を入力とし、図１６のプラン生成ルールを用いた場合の検索プラン生成部３３の出力結果例である。

【０１６１】

本プラン生成を行うに当たり、構造化文書データベースへの前提として以下のものを想定している。

・要素名称生起インデックスファイルの中に「特許」というキーが存在している。

・データ生起インデックスファイルの中に「検索」というキーが存在している。

【０１６２】

検索グラフの全Ｇノードを候補Ｇノードリスト３３４に登録した後、図１８のフローチャートにしたがってシミュレーションする。ステップＳ３からステップＳ７の１処理を１サイクルとして、変数ｒｓｅｔの変化を追ってみると、次のようになる。

【０１６３】

（第１サイクル）

ｒｓｅｔ＝｛＜ＴＡＧ０１、ルール０１、１．０＞、＜ＴＡＧ０３、ルール０３、０．２＞、＜ＣＭＰ０１、ルール３１、１．０＞、＜ＣＭＰ０１、ルール３２、０．１＞｝

ここで、＜ＣＭＰ０１、ルール３２、０．１＞が選択され、ＦＩＮＤが出力される。

伝播するＧノード群は｛＄ｘ１｝である。

（第２サイクル）

ｒｓｅｔ＝｛＜ＴＡＧ０１、ルール０１、１．０＞、＜ＴＡＧ０３、ルール０３、０．２＞｝

ここで、＜ＴＡＧ０３、ルール０３、０．２＞が選択され、ＰＡＴＨＥＸＰＡＮＤ２が出力される。

伝播するＧノード群は｛＄＿２、＄＿３｝である。

【０１６４】

（第３サイクル）

ｒｓｅｔ＝｛＜ＴＡＧ０１、ルール０１、１．０＞、＜ＴＡＧ０４、ルール０

２、０．５＞、＜ＴＡＧ０２、ルール０６、０．６＞}

ここで、＜ＴＡＧ０４、ルール０２、０．３＞が選択され、ＰＡＴＨＥＸＰＡＮＤ２が出力される。

伝播するＧノード群は｛＄＿４｝である。

【０１６５】

（第４サイクル）

ｒｓｅｔ＝｛＜ＴＡＧ０１、ルール０１、１．０＞、＜ＶＡＬ０１、ルール２１、０．２＞、＜ＴＡＧ０２、ルール０６、０．６＞}

ここで、＜ＶＡＬ０１、ルール２１、０．２＞が選択され、ＶＡＬＵＥが出力される。

伝播するＧノード群は｛＄ｘ１｝である。

【０１６６】

（第５サイクル）

ｒｓｅｔ＝｛＜ＴＡＧ０１、ルール０１、１．０＞、＜＄ｘ１、ルール１１、０．５＞、＜ＴＡＧ０２、ルール０６、０．６＞}

ここで、＜＄ｘ１、ルール１１、０．５＞が選択され、ＪＯＩＮが出力される。

伝播するＧノード群は｛｝である。

【０１６７】

（第６サイクル）

ｒｓｅｔ＝｛＜ＴＡＧ０１、ルール０１、１．０＞、＜ＴＡＧ０２、ルール０６、０．６＞}

ここで、＜ＴＡＧ０２、ルール０６、０．６＞が選択され、ＮＯＰが出力される。

伝播するＧノード群は｛｝である。

【０１６８】

（第７サイクル）

ｒｓｅｔ＝｛＜ＣＯＮ０１、ルール７１、１．０＞}

ここで、＜ＣＯＮ０１、ルール７１、１．０＞が選択され、ＣＯＮＳＴＲＵＣ

Ｔが出力される。

【０１６９】

この実行プランリスト３３５の意味は、以下のようなものである。

【０１７０】

（ステップ１）

「検索」という文字列データを含むＤノード群を検索する。データ生起インデックスファイルには「検索」というキーが存在しているため、この情報を優先的に利用する。

（ステップ２）

「特許」要素名称を持つ子供Ｄノード群を取り出す。要素名称生起インデックスファイルの中に「特許」というキーが存在しているため、この情報を優先的に利用する。

（ステップ３）

上記Ｄノード群で「名称」要素名称を持つ子供Ｄノード群を取り出す。

（ステップ４）

上記Ｄノード群で要素データを持つ子供Ｄノード群を取り出す。

（ステップ５）

ステップ１で検索したＤノード群とステップ４で取り出したＤノード群の結合（ＪＯＩＮ）を取る。

（ステップ６）

「特許」より上位の文書パスは「ｒｏｏｔ／＊」なので、何もしない。

（ステップ７）

上記Ｄノード群のデータを使って「文献」情報を作り出す。

【０１７１】

図２２に、図２１で示した検索プランの実行イメージを示す。

【０１７２】

ステップ１において、図１４のデータ生起インデックスファイルに「検索」というキーが存在するので、直ちに、Ｄノード群が得られる。

【０１７３】

一方、ステップ２において、図１３の要素名称生起インデックスファイルに「特許」というキーが存在するので、直ちに、＄２と＄３が具体化される。そして、ステップ３において、図５から＄４が具体化される。そして、ステップ４において、要素データを持つＤノード群を取り出す。ステップ５において、ステップ１で検索したＤノード群とステップ４で取り出したＤノード群の結合（ＪＯＩＮ）を取る。

【０１７４】

このように、インデックスを用いて効率的に検索プランを生成していることがわかる。

【０１７５】

図２３に、図１９の検索要求を処理した検索結果の一例を示す。

【０１７６】

検索結果もＸＭＬで表現され、「文献」情報のリストとして表示されている。「情報検索装置」など「検索」という文字列を含んでいる。

【０１７７】

ここで、比較のために従来手法でアプローチした場合について説明する。

【０１７８】

図２４に、従来手法でアプローチした場合の検索プランの一例を示す。

【０１７９】

この従来手法は、ルート要素から始まって、親要素から子供要素群へと展開して、階層的な構造上の構成要素を特定する検索処理方式である。

【０１８０】

この実行プランリストの意味は、以下のようなものである。

【０１８１】

（ステップ１）

ｒｏｏｔに相当するＤノード群を取り出す。

（ステップ２）

上記Ｄノード群の子孫Ｄノード群を取り出す。

（ステップ３）

上記Ｄノード群で「特許」要素名称を持つ子供Ｄノード群を取り出す。

（ステップ４）

上記Ｄノード群で「名称」要素名称を持つ子供Ｄノード群を取り出す。

（ステップ５）

上記Ｄノード群で要素データを持つ子供Ｄノード群を取り出す。

（ステップ６）

上記Ｄノード群でデータが「検索」という文字列データを含むＤノード群を選択する。

（ステップ７）

上記Ｄノード群のデータを使って「文献」情報を作り出す。

【０１８２】

このように、登録された文書数が増大し階層木の深さと幅が増大すると、検索処理に要する計算量が膨大なものになってしまう。

【０１８３】

図２５に、図２４で示した従来手法でアプローチした場合の検索プランの実行イメージを示す。

【０１８４】

ステップ２での階層木の展開コストが膨大になってしまうことが容易に想像される。

【０１８５】

以下では、本実施形態における検索結果をＧＵＩ（グラフィカル・ユーザ・インタフェース）的に表示する例について説明する。

【０１８６】

図２６に、図１０で示された図９の検索結果をデータ表示用のフィルタープログラムを通してＧＵＩ的に表示した一例を示す。

【０１８７】

「特許」情報に対して、概念「情報モデル」での分類と概念「情報操作」での分類の２分類軸を設定して、「出願番号」と「情報モデル」軸と「情報操作」軸を抽出して「文献」情報として検索した結果であるが、「情報モデル」軸を横軸

に、「情報操作」軸に設定して、２軸のクロスしたデータが「出願番号」情報である。

【０１８８】

ＸＭＬにはスタイルシートという表示フォーマットがあり、ＸＭＬドキュメントをＷＷＷブラウザに表示したり、プリンタから印刷したりするときに用いる。スタイルシートの言語として、ＸＳＬ（Ｅｘｔｅｎｓｉｂｌｅ　Ｓｔｙｌｅ　Ｌａｎｇｕａｇｅ）が標準の規約として用意されており、これを用いれば図２６に示すような情報を出力することができる。

【０１８９】

図２７に、図１０で示された図９の検索結果をデータ表示用のフィルタープログラムを通してＧＵＩ的に表示した他の例を示す。

【０１９０】

特許の出願件数を年度別に折れ線グラフ表示させたもので、これも同様に、『特許の出願件数を年度別に集計する』検索要求を処理した結果であるＸＭＬデータに対してスタイルシートを適用すれば、図２７に示すような情報を出力することができる。

【０１９１】

ここで、図１の構造化文書データベース・システムの実現方法のバリエーションについて説明する。

【０１９２】

本システムは、インターネットもしくはＬＡＮなどのネットワークを介して他の計算機から検索要求を受け付け、検索を実行し、ネットワークを介して当該他の計算機に検索結果を返すように実現することも可能である。

【０１９３】

この場合、他の計算機から図７～図９のような検索要求を受ける代わりに、他の計算機において構文解析し図１１／図１２のような検索グラフを作成し、これを受けるようにしてもよい。あるいは、図７～図９のような検索要求と図１１／図１２のような検索グラフのいずれによっても受け付け可能としてもよい。

【０１９４】

また、要求制御部１、格納処理部２、検索処理部３を１台の計算機上に実装してもよいし、２台または３台の計算機上に別々に実装してもよい。

【０１９５】

また、要求制御部１、格納処理部２、検索処理部３のそれぞれを実現するプログラムは、記録媒体または通信媒体によって受け渡すことが可能である。この場合、要求制御部１、格納処理部２、検索処理部３の全てを実現するプログラムを１つまたは１組の記録媒体に格納して受け渡しすることも、要求制御部１、格納処理部２、検索処理部３の一部のみを実現するプログラムを１つまたは１組の記録媒体に格納して受け渡しすることも可能である。

【０１９６】

また、例えば、検索処理部３を含むシステムと、格納処理部２とデータファイル４とインデックスファイル５を含むシステムとが、互いに独立したシステムであってもよい。また、検索処理部３を含むシステムをサーバとして構成してもよいし、各クライアントに搭載するようにしてもよい。

【０１９７】

もちろん、本システムは、１つのスタンドアローンのシステムとして実現可能である。

【０１９８】

なお、以上の各機能は、ソフトウェアとしても実現可能である。

【０１９９】

また、本実施形態は、コンピュータに所定の手段を実行させるための（あるいはコンピュータを所定の手段として機能させるための、あるいはコンピュータに所定の機能を実現させるための）プログラムを記録したコンピュータ読取り可能な記録媒体としても実施することもできる。

【０２００】

本発明は、上述した実施の形態に限定されるものではなく、その技術的範囲において種々変形して実施することができる。

【０２０１】

【発明の効果】

本発明によれば、構造化文書データベースに関する情報を有効に利用しながら、検索要求から生成した検索グラフを最適に巡回することで最適な検索プランを生成し実行することによって、計算量を増大させずに、（曖昧パスを含む）文書が持つ階層構造に対する多様な検索指定による検索を可能にすることができる。

【図面の簡単な説明】

【図１】

本発明をの一実施形態に係る構造化文書データベース・システムのシステム構成例を示す図

【図２】

構造化文書の一例を示す図

【図３】

概念情報の一例を示す図

【図４】

概念情報の一例を示す図

【図５】

構造化文書データベースの概念的な構造例を示する図

【図６】

構造化文書データベースへの構造化文書の格納コマンドの一例を示す図

【図７】

構造化文書データベースへの検索コマンドの一例を示す図

【図８】

構造化文書データベースへの検索コマンドの他の例を示す図

【図９】

構造化文書データベースへの検索コマンドのさらに他の例を示す図

【図１０】

検索要求を処理した検索結果の一例を示す図

【図１１】

検索要求に対して検索グラフ生成部が生成する検索グラフの一例を示す図

【図１２】

検索要求に対して検索グラフ生成部が生成する検索グラフの一例を示す図

【図１３】

インデックスファイルの一種である要素名称生起インデックスの概念的な構造例を示す図

【図１４】

インデックスファイルの一種であるデータ生起インデックスの概念的な構造例を示す図

【図１５】

検索プラン生成部の構成例を示す図

【図１６】

検索プラン生成部で利用するプラン生成ルールの一例を示す図

【図１７】

検索プラン実行部で利用するオペレータの一例を示す図

【図１８】

検索プラン生成部の処理手順の一例を示すフローチャート

【図１９】

構造化文書データベースへの検索コマンドのさらに他の例を示す図

【図２０】

検索要求に対して検索グラフ生成部が生成する検索グラフの他の例を示す図

【図２１】

検索プラン生成部により生成された検索プランの一例を示す図

【図２２】

検索プランの実行イメージを示す図

【図２３】

検索要求を処理した検索結果の他の例を示す図

【図２４】

従来手法でアプローチした場合の検索プランを示す図

【図２５】

従来手法でアプローチした場合の検索プランの実行イメージを示す図

【図２６】

検索結果をデータ表示用のフィルタープログラムを通してＧＵＩ的に表示した一例を示す図

【図２７】

検索結果をデータ表示用のフィルタープログラムを通してＧＵＩ的に表示した他の例を示す図

【符号の説明】

１…要求制御部

２…格納処理部

３…検索処理部

４…データファイル

５…インデックスファイル

６…プラン生成ルール

１１…要求受付部

１２…要求処理部

１３…結果処理部

２１…格納要求構文解析部

２２…データ格納部

２３…インデックス格納部

３１…検索要求構文解析部

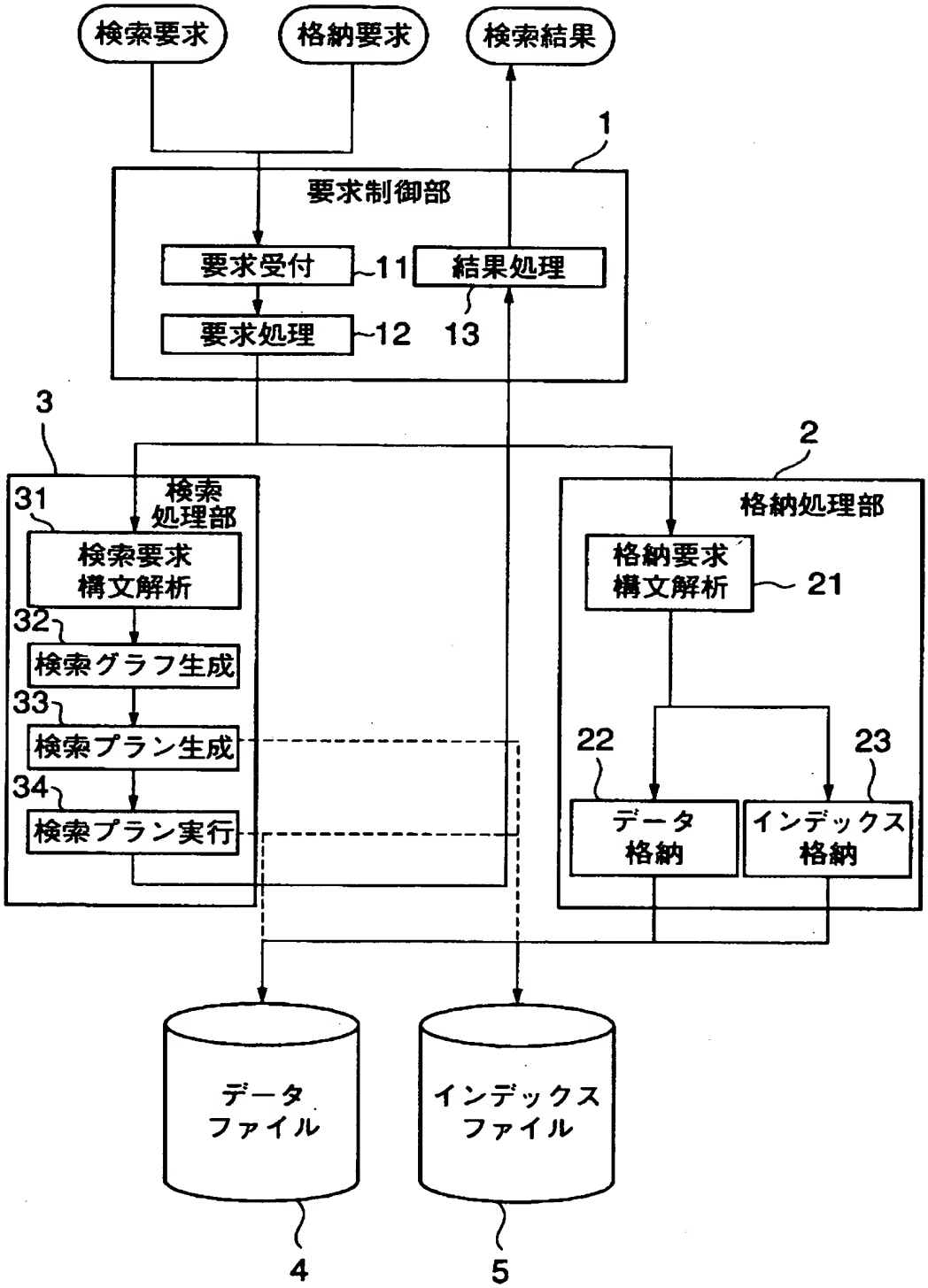
３２…検索グラフ生成部

３３…検索プラン生成部

３４…検索プラン実行部

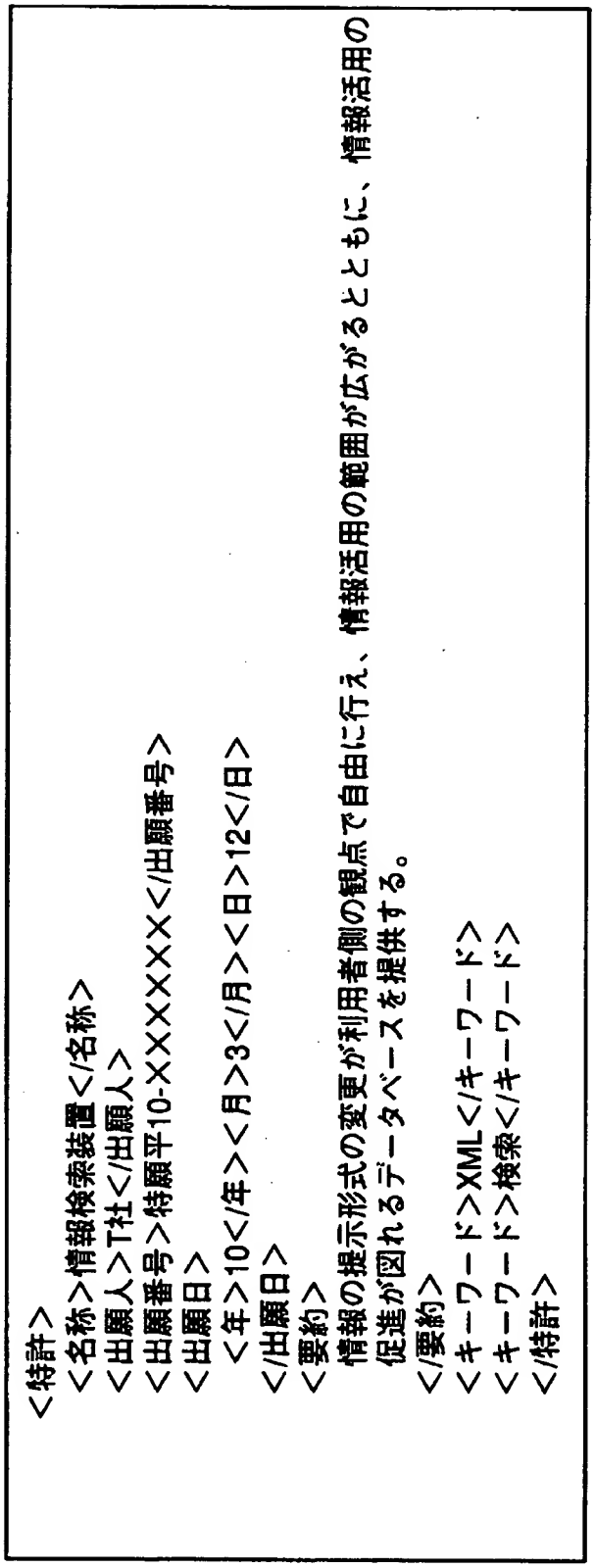
【書類名】　図面

【図１】



出証特２０００－３０８５７００

【図２】


<特許>
<名称>情報検索装置</名称>
<出願人>T社</出願人>
<出願番号>特願平10-XXXXXXX</出願番号>
<出願日>
<年>10</年><月>3</月><日>12</日>
</出願日>
<要約>
情報の提示形式の変更が利用者側の観点で自由に行え、情報活用の範囲が広がるとともに、情報活用の促進が図れるデータベースを提供する。
</要約>
<キーワード>XML</キーワード>
<キーワード>検索</キーワード>
</特許>

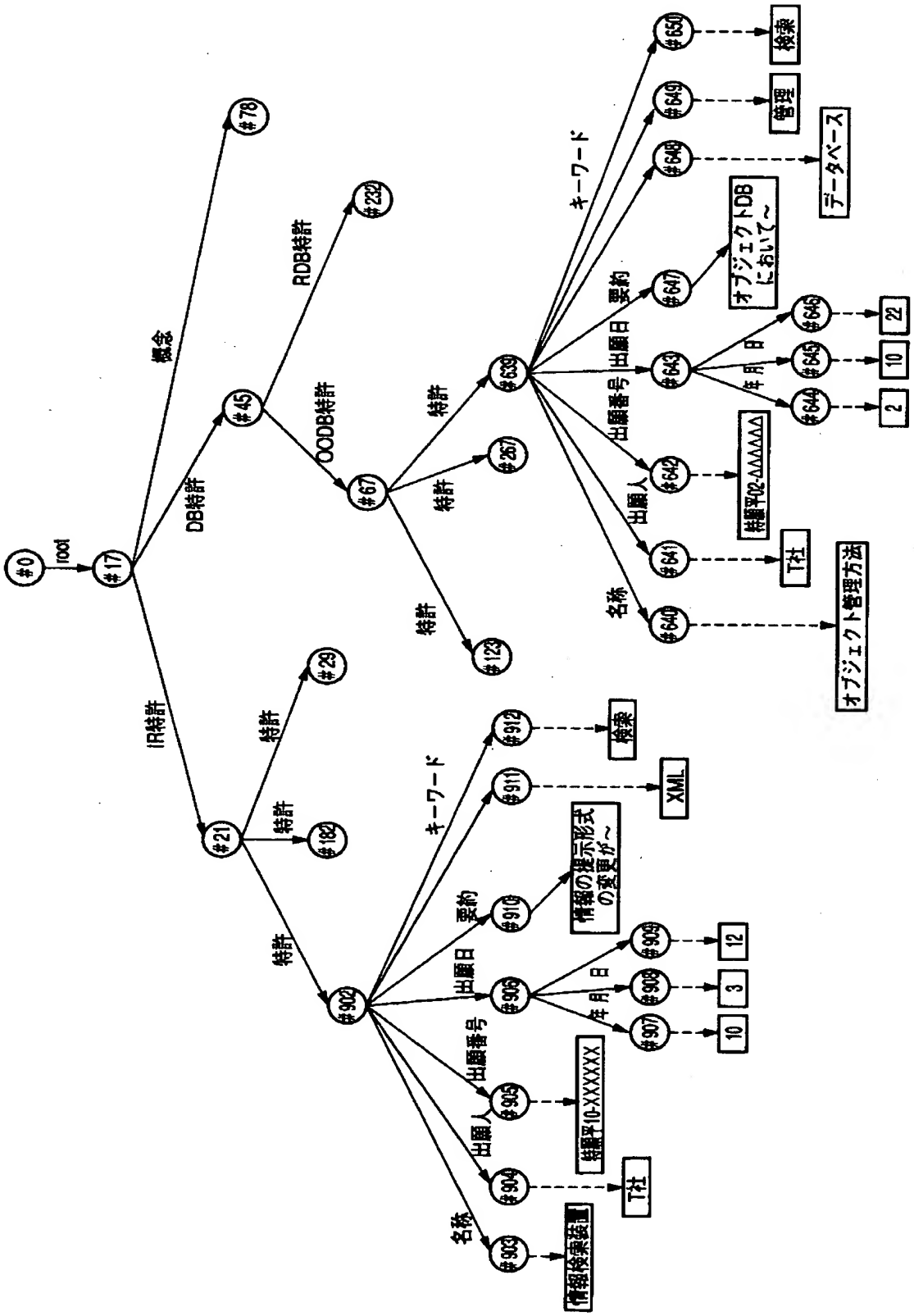
【図３】

```
<概念名前＝“情報モデル”>
  <概念名前＝“ドキュメント”>
    <概念名前＝“構造化ドキュメント”>
      <概念名前＝“XML”/>
      <概念名前＝“SGML”/>
    </概念>
    <概念名前＝“非構造ドキュメント”>
      <概念名前＝“テキスト”/>
    </概念>
  </概念>
  <概念名前＝“リレーション”>
    ……
  </概念>
  <概念名前＝“オブジェクト”>
    ……
  </概念>
</概念>
```

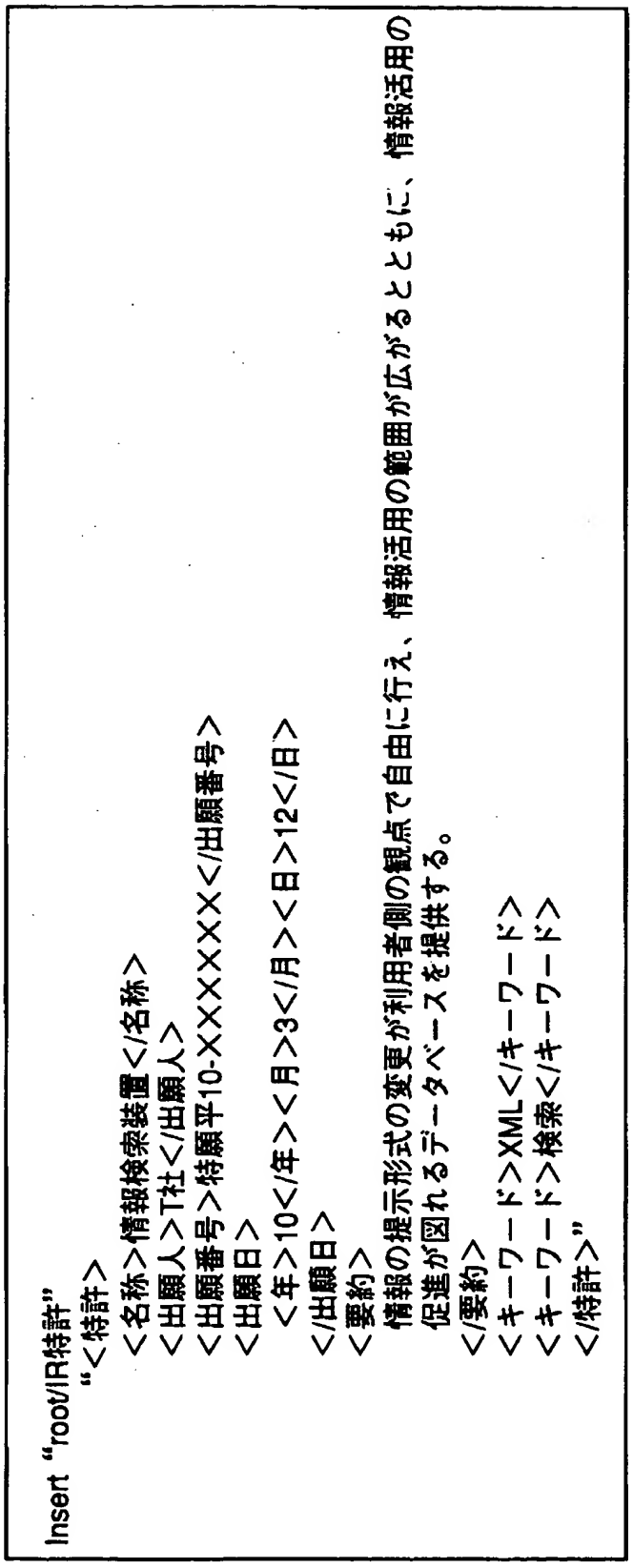
【図４】

```
<概念名前＝“情報操作”>
  <概念名前＝“検索”/>
  <概念名前＝“格納”/>
  <概念名前＝“加工”/>
  <概念名前＝“流通”/>
</概念>
```

【図５】


#0
root
#17
概念
#78
DB特許
#45
RDB特許
#232
OODB特許
#67
特許
#639
キーワード
#650
検索
#649
管理
#648
データベース
要約
#647
オブジェクトDBにおいて～
出願日
#643
年月日
#646
22
#645
10
#644
2
出願番号
#642
特願平02-AAAAAA
出願人
#641
T社
名称
#640
オブジェクト管理方法
特許
#267
特許
#123
IR特許
#21
特許
#29
特許
#182
特許
#902
キーワード
#912
検索
#911
XML
要約
#910
情報の提示形式の変更が～
出願日
#909
年月日
#909
12
#908
3
#907
10
出願番号
#905
特願平10-XXXXXX
出願人
#904
T社
名称
#903
情報検索装置

【図６】


Insert “root/IR特許”
“<特許>
<名称>情報検索装置</名称>
<出願人>T社</出願人>
<出願番号>特願平10-XXXXXX</出願番号>
<出願日>
<年>10</年><月>3</月><日>12</日>
</出願日>
<要約>
情報の提示形式の変更が利用者側の観点で自由に行え、情報活用の範囲が広がるとともに、情報活用の促進が図れるデータベースを提供する。
</要約>
<キーワード>XML</キーワード>
<キーワード>検索</キーワード>
</特許>”

【図７】

```
Select
  <文献>
    <出願番号>$x1</出願番号>
  </文献>
Where
  <*/特許>
    <出願番号>$x1</出願番号>
    <キーワード>$x2</キーワード>
  </特許>From “root/”
  $x2= “検索”
```

【図８】

```
Select
  <文献>
    <出願番号>$x1</出願番号>
  </文献>
Where
  <*/特許>
    <出願番号>$x1</出願番号>
    <キーワード>$x2</キーワード>
  </特許>From “root/”
  <概念名前＝“ドキュメント”>
    <*/概念名前＝$x2/>
  </概念>From “root/”
```

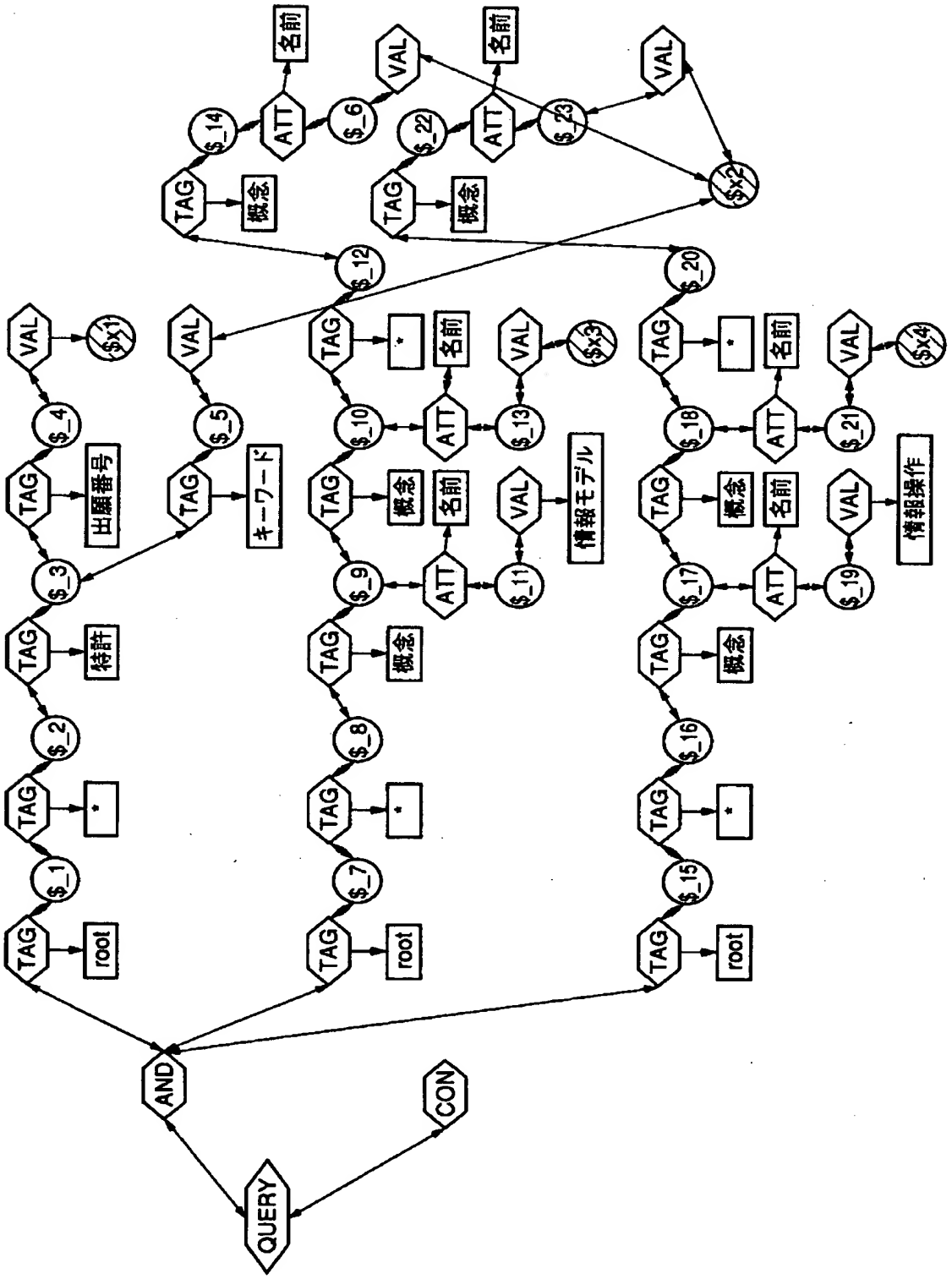
【図９】

```
Select
  <文献>
    <出願番号>$x1</出願番号>
    <分類軸＝“情報モデル”>$3</分類>
    <分類軸＝“情報操作”>$4</分類>
  </文献>
Where
  <*/特許>
    <出願番号>$x1</出願番号>
    <キーワード>$x2</キーワード>
  </特許>From “root/”
  <*/概念名前＝“情報モデル”>
    <概念名前＝$3>
      <*/概念名前＝$x2/>
    </概念>
  </概念>From “root/”
  <*/概念名前＝“情報操作”>
    <概念名前＝$4>
      <*/概念名前＝$x2/>
    </概念>
  </概念>From “root/”
```

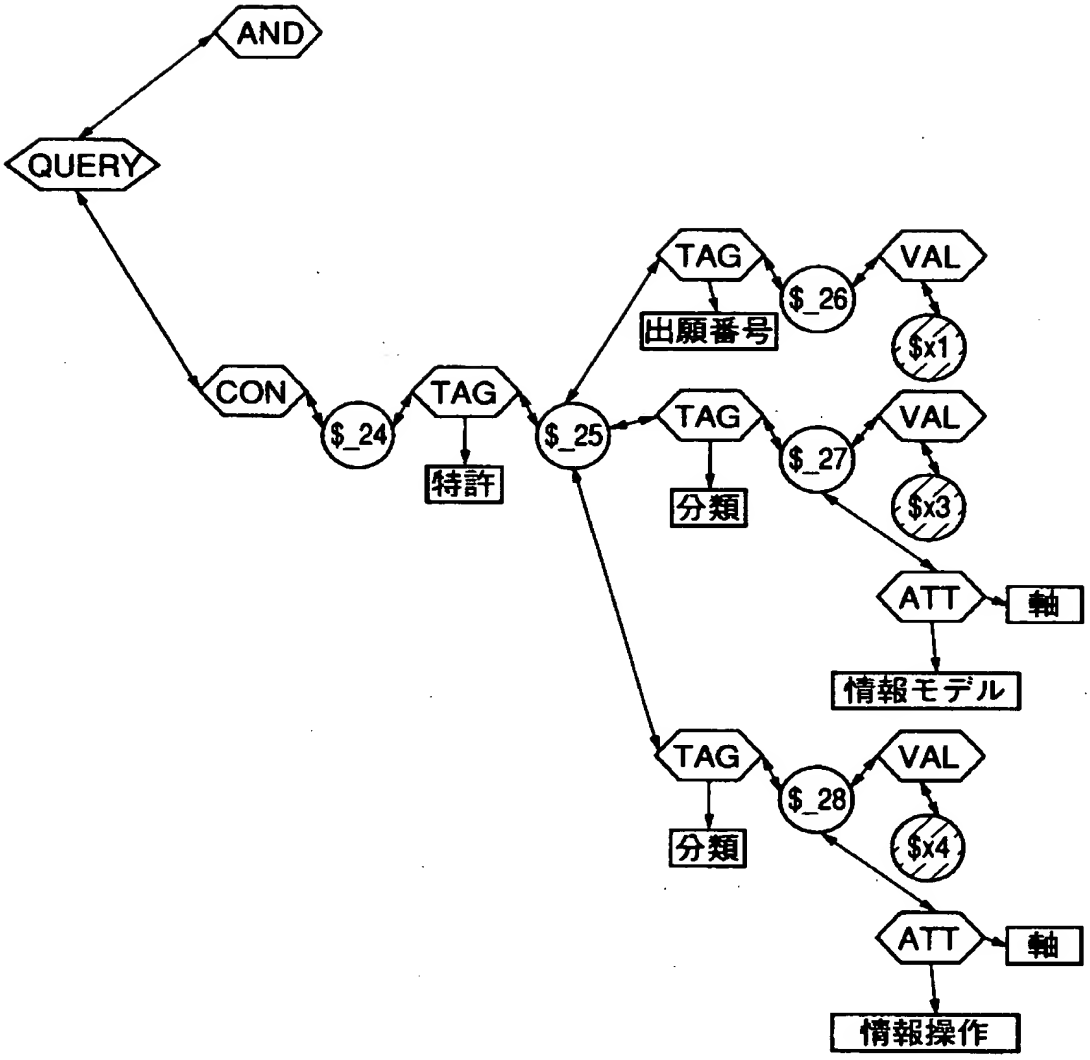
【図１０】

```
<結果>
  <文献>
    <出願番号>特願平10-XXXXXX</出願番号>
    <分類軸＝“情報モデル”>ドキュメント</分類>
    <分類軸＝“情報操作”>検索</分類>
  </文献>
  <文献>
    <出願番号>特願平09-□□□□□□</出願番号>
    <分類軸＝“情報モデル”>リレーション</分類>
    <分類軸＝“情報操作”>格納</分類>
  </文献>
  <文献>
    <出願番号>特願平10-000000</出願番号>
    <分類軸＝“情報モデル”>リレーション</分類>
    <分類軸＝“情報操作”>検索</分類>
  </文献>
</結果>
```

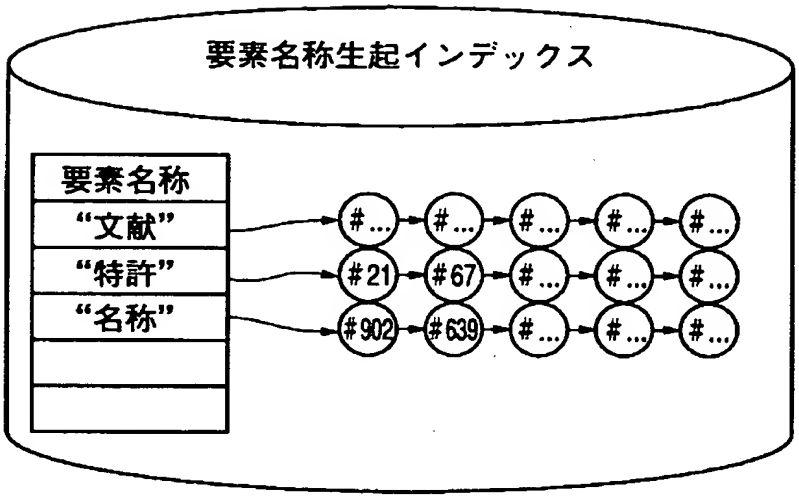
【図１１】


QUERY
AND
CON
TAG
root
*
概念
特許
出願番号
キーワード
情報モデル
情報操作
名前
VAL
ATT
$_1
$_2
$_3
$_4
$_5
$_6
$_7
$_8
$_9
$_10
$_11
$_12
$_13
$_14
$_15
$_16
$_17
$_18
$_19
$_20
$_21
$_22
$_23
$x1
$x2
$x3
$x4

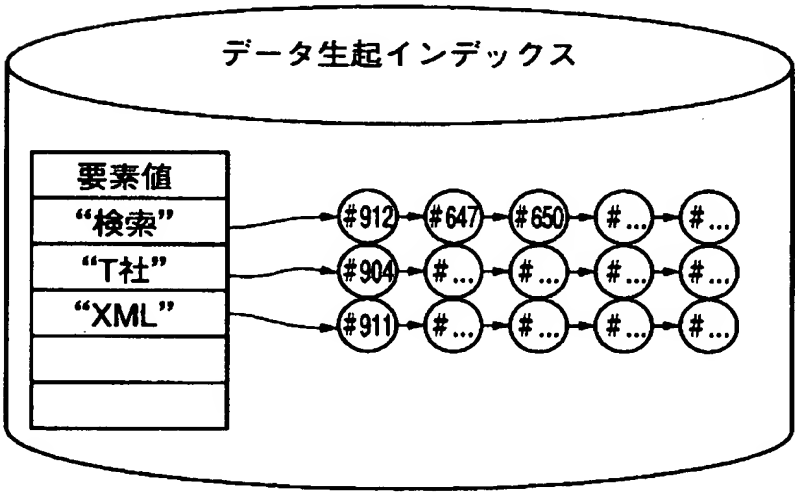
【図１２】


AND
QUERY
CON
$_24
TAG
特許
$_25
TAG
出願番号
$_26
VAL
$x1
TAG
分類
$_27
VAL
$x3
ATT
軸
情報モデル
TAG
分類
$_28
VAL
$x4
ATT
軸
情報操作

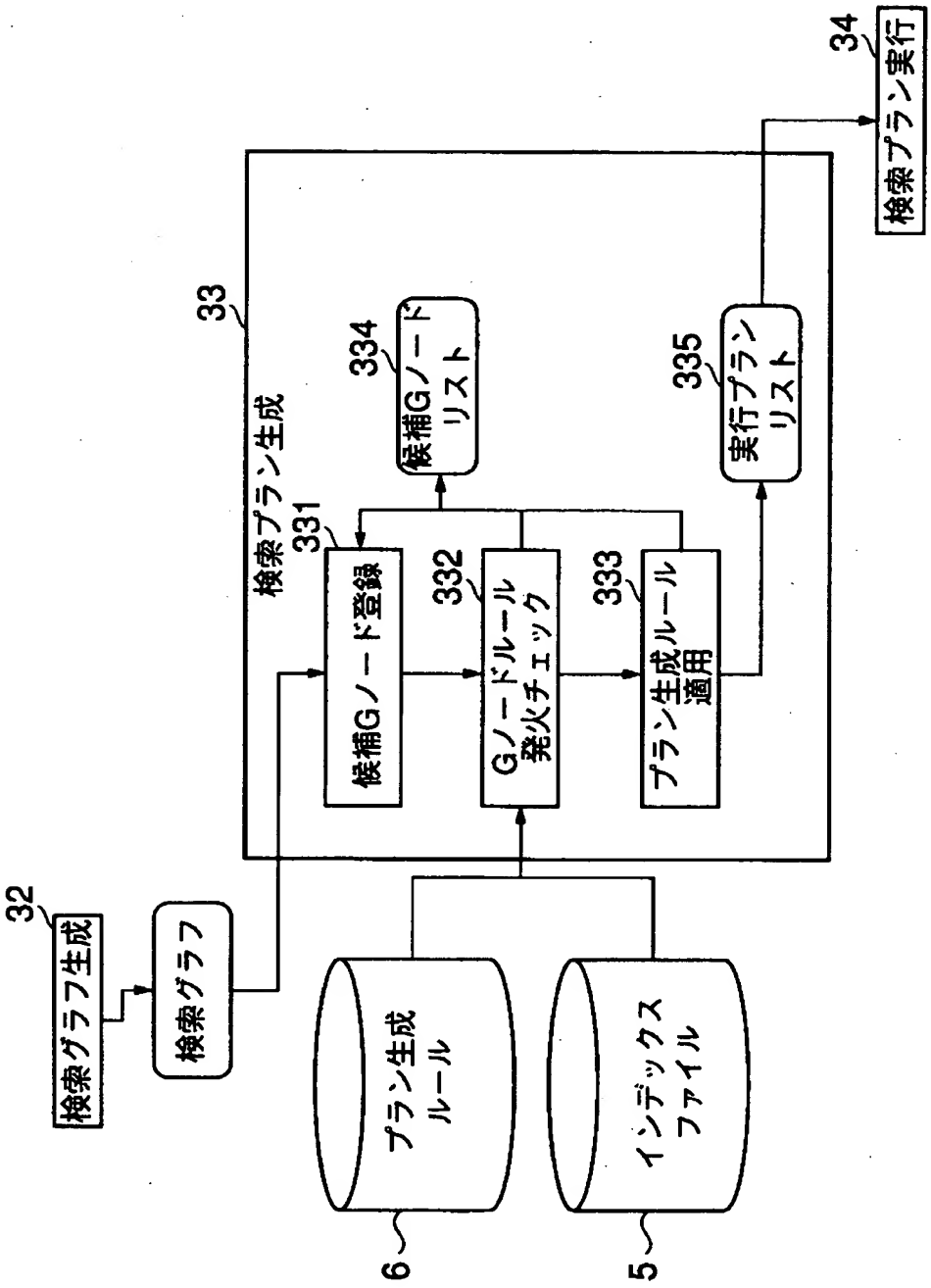
【図１３】


要素名称生起インデックス
要素名称
“文献”
“特許”
“名称”
#...
#21
#67
#902
#639

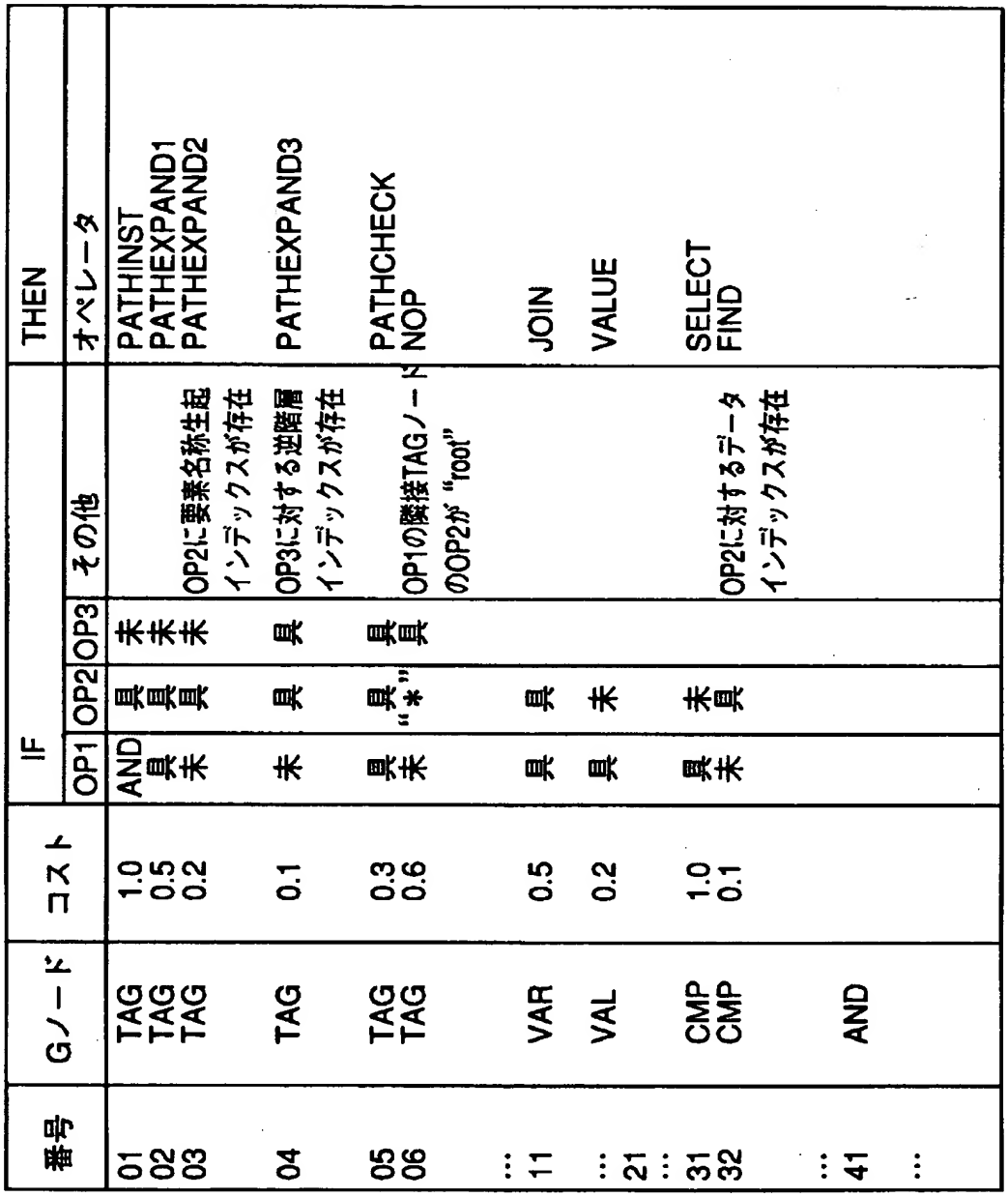
【図１４】


データ生起インデックス
要素値
“検索”
“T社”
“XML”
#912
#647
#650
#904
#911
#...

【図１５】

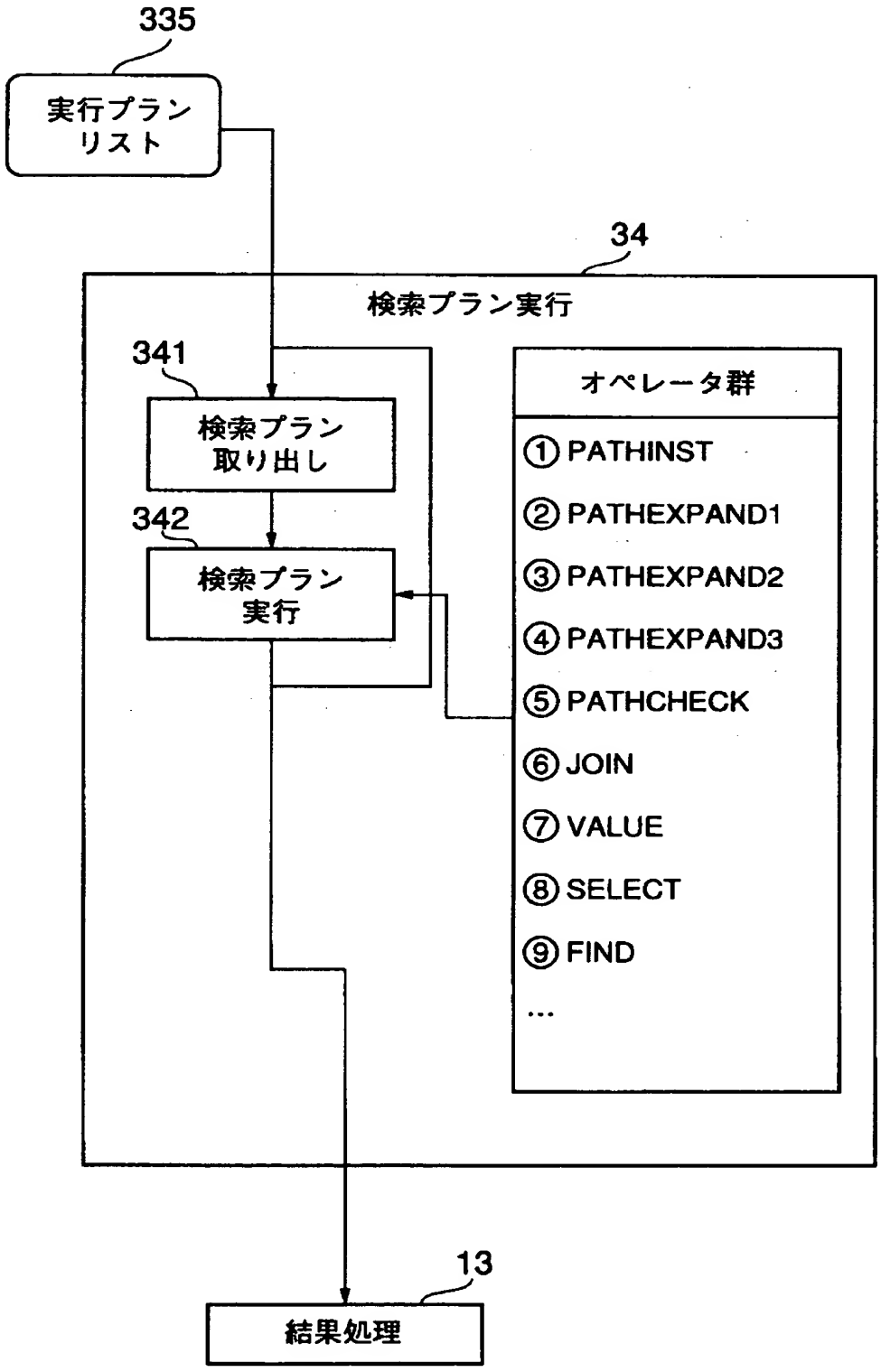

32
検索グラフ生成
検索グラフ
33
検索プラン生成
331
候補Gノード登録
332
Gノードルール発火チェック
333
プラン生成ルール適用
334
候補Gノードリスト
335
実行プランリスト
34
検索プラン実行
6
プラン生成ルール
5
インデックスファイル

【図１６】

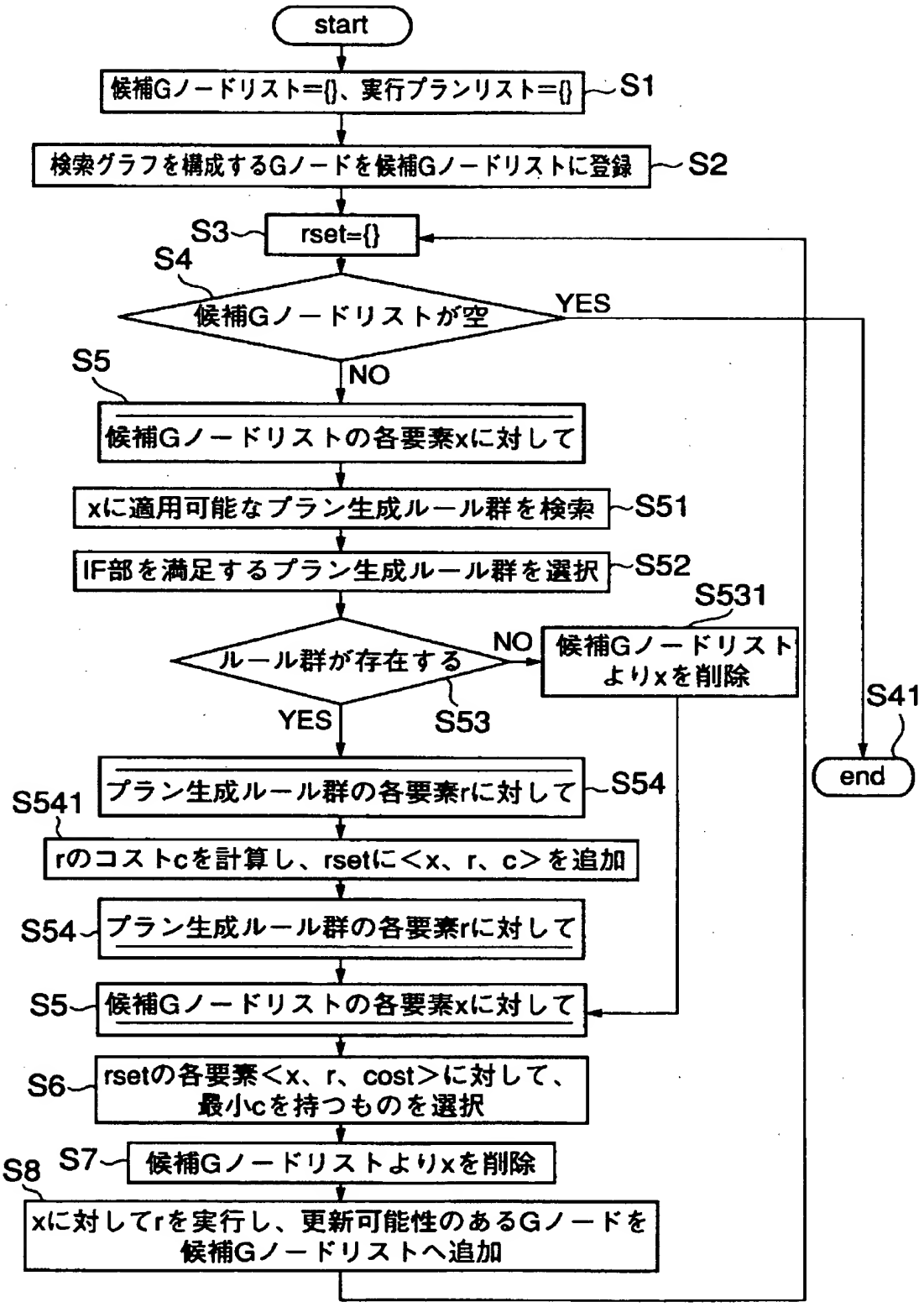


番号	Gノード	コスト	IF				THEN
			OP1	OP2	OP3	その他	オペレータ
01	TAG	1.0	AND	具	未		PATHINST
02	TAG	0.5	具	具	未		PATHEXPAND1
03	TAG	0.2	未	具	未	OP2に要素名称生起インデックスが存在	PATHEXPAND2
04	TAG	0.1	未	具	具	OP3に対する逆階層インデックスが存在	PATHEXPAND3
05	TAG	0.3	具	具	具		PATHCHECK
06	TAG	0.6	未	“*”	具	OP1の隣接TAGノードのOP2が“root”	NOP
…							
11	VAR	0.5	具	具			JOIN
…							
21	VAL	0.2	具	未			VALUE
…							
31	CMP	1.0	具	未			SELECT
32	CMP	0.1	未	具		OP2に対するデータインデックスが存在	FIND
…							
41	AND						
…							

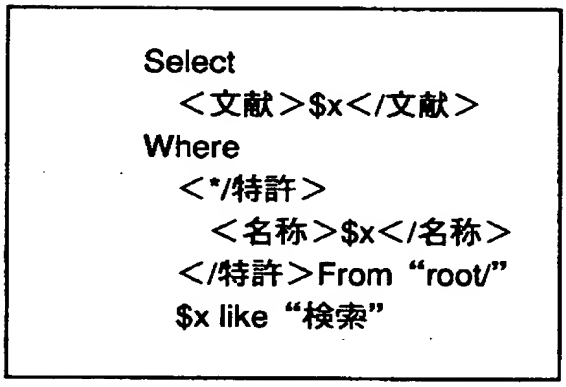
【図１７】


335
実行プラン
リスト
34
検索プラン実行
341
検索プラン
取り出し
342
検索プラン
実行
オペレータ群
① PATHINST
② PATHEXPAND1
③ PATHEXPAND2
④ PATHEXPAND3
⑤ PATHCHECK
⑥ JOIN
⑦ VALUE
⑧ SELECT
⑨ FIND
…
13
結果処理

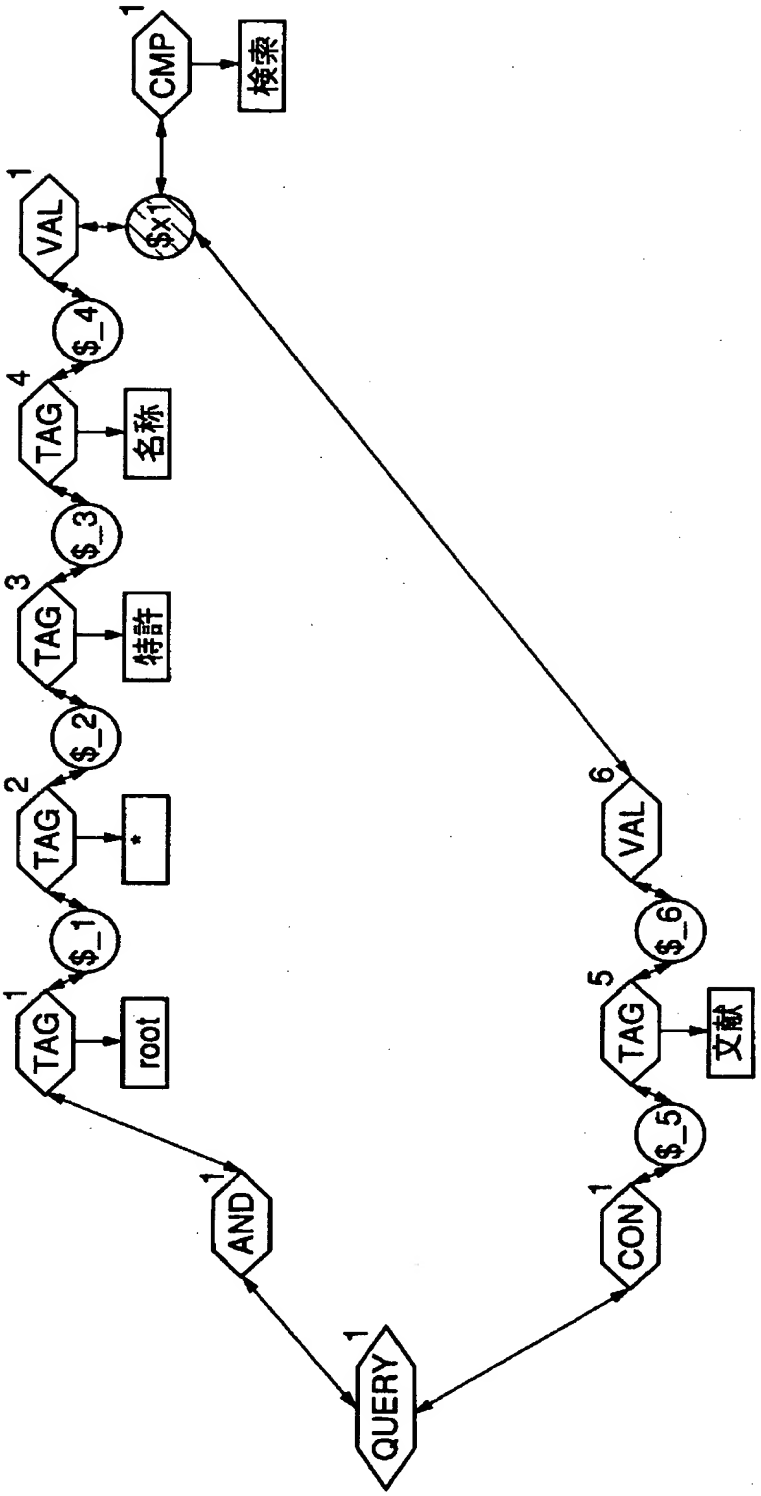
【図１８】


start
候補Gノードリスト={}、実行プランリスト={}
S1
検索グラフを構成するGノードを候補Gノードリストに登録
S2
S3
rset={}
S4
候補Gノードリストが空
YES
NO
S5
候補Gノードリストの各要素xに対して
xに適用可能なプラン生成ルール群を検索
S51
IF部を満足するプラン生成ルール群を選択
S52
S531
ルール群が存在する
NO
候補Gノードリストよりxを削除
YES
S53
S41
end
S541
プラン生成ルール群の各要素rに対して
S54
rのコストcを計算し、rsetに<x、r、c>を追加
S54
プラン生成ルール群の各要素rに対して
S5
候補Gノードリストの各要素xに対して
S6
rsetの各要素<x、r、cost>に対して、最小cを持つものを選択
S8
S7
候補Gノードリストよりxを削除
xに対してrを実行し、更新可能性のあるGノードを候補Gノードリストへ追加

【図１９】



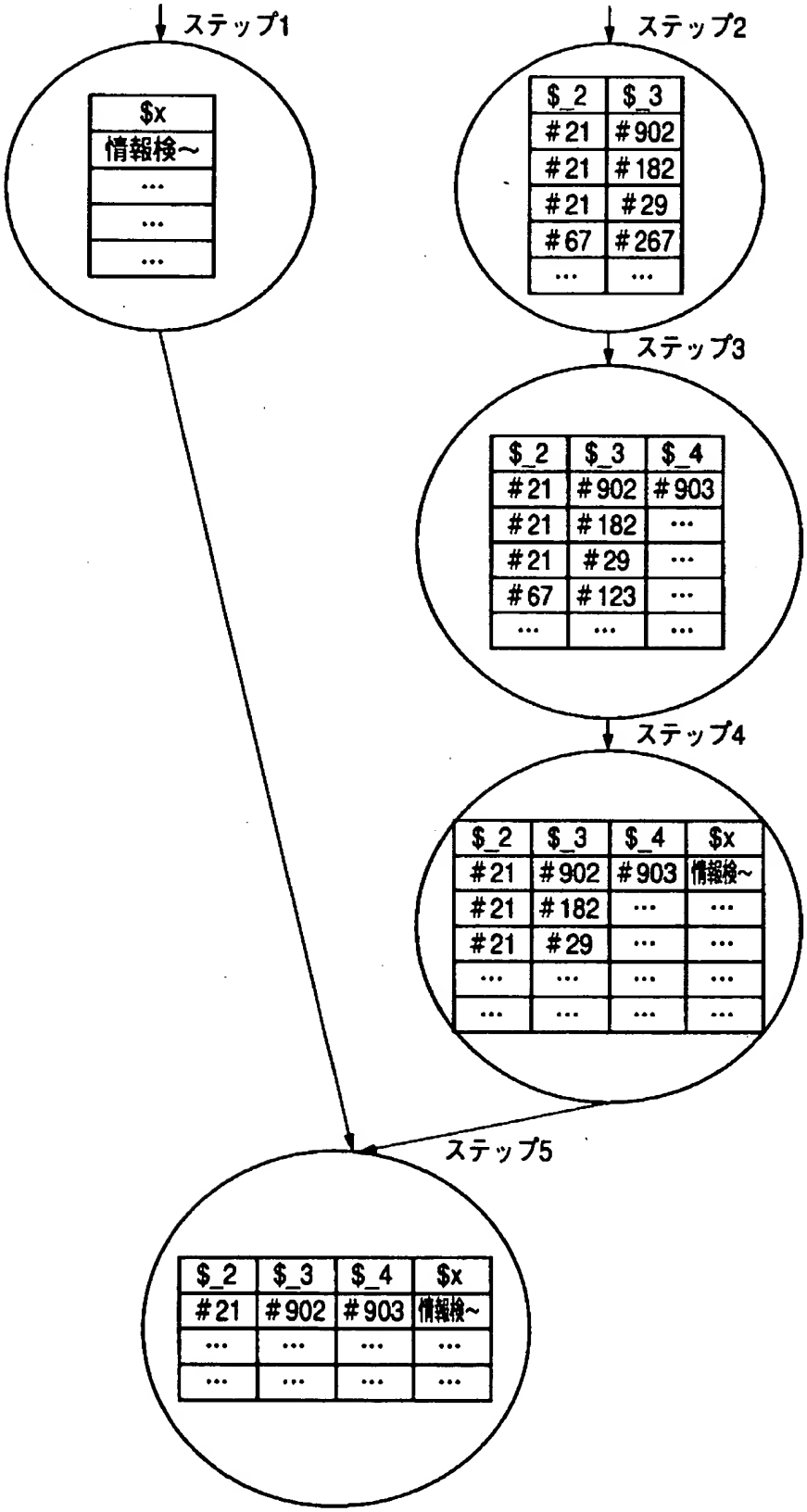
【図２０】


QUERY
AND
CON
TAG
root
TAG
*
TAG
特許
TAG
名称
VAL
TAG
文献
VAL
CMP
検索
$X1
$_1
$_2
$_3
$_4
$_5
$_6

【図２１】

ステップ	Gノード	ルール番号	オペレータ
ステップ1	CMP	ルール32適用	FIND
ステップ2	TAG	ルール03適用	PATHEXPAND2
ステップ3	TAG	ルール02適用	PATHEXPAND1
ステップ4	VAL	ルール21適用	VALUE
ステップ5	VAR	ルール11適用	JOIN
ステップ6	TAG	ルール06適用	NOP
ステップ7	CON	ルール適用	CONSTRUCT

【図２２】


ステップ1
$x
情報検~
ステップ2
$_2
$_3
#21
#902
#21
#182
#21
#29
#67
#267
ステップ3
$_2
$_3
$_4
#21
#902
#903
#21
#182
#21
#29
#67
#123
ステップ4
$_2
$_3
$_4
$x
#21
#902
#903
情報検~
#21
#182
#21
#29
ステップ5
$_2
$_3
$_4
$x
#21
#902
#903
情報検~

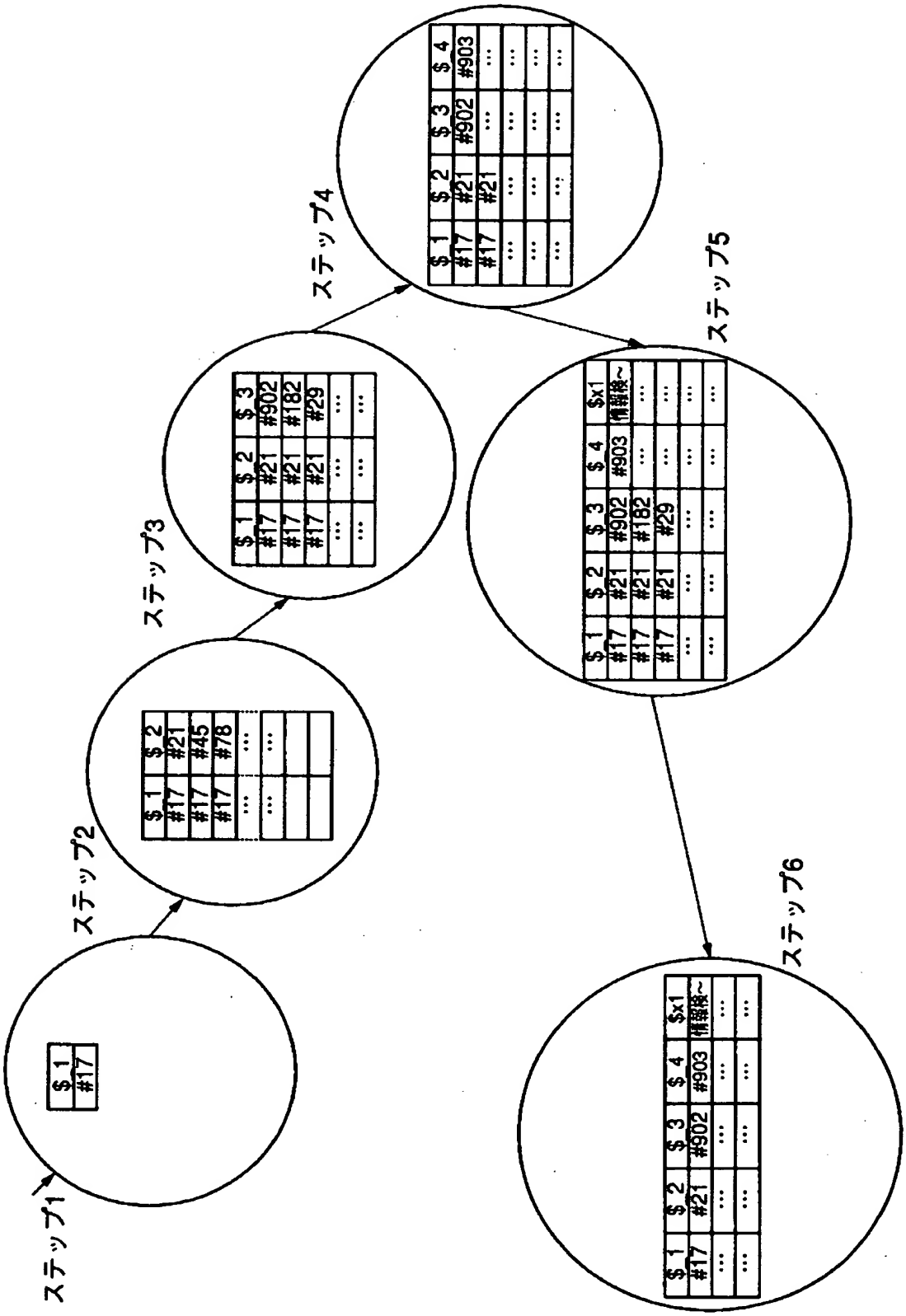
【図２３】

```
<結果>
  <文献>情報検索装置</文献>
  <文献>データ検索方式</文献>
  <文献>検索処理の最適化の方法と検索装置</文献>
</結果>
```

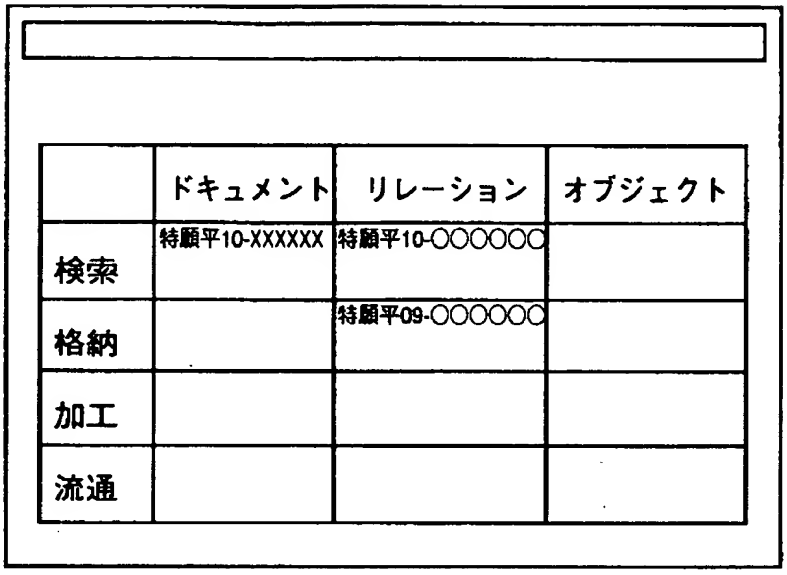
【図２４】

ステップ	Gノード	ルール番号	オペレータ
ステップ1	TAG	ルール01適用	PATHINST
ステップ2	TAG	ルール02適用	PATHEXPAND1
ステップ3	TAG	ルール02適用	PATHEXPAND1
ステップ4	TAG	ルール02適用	PATHEXPAND1
ステップ5	VAL	ルール21適用	VALUE
ステップ6	CMP	ルール31適用	SELECT
ステップ7	CON	ルール適用	CONSTRUCT

【図２５】

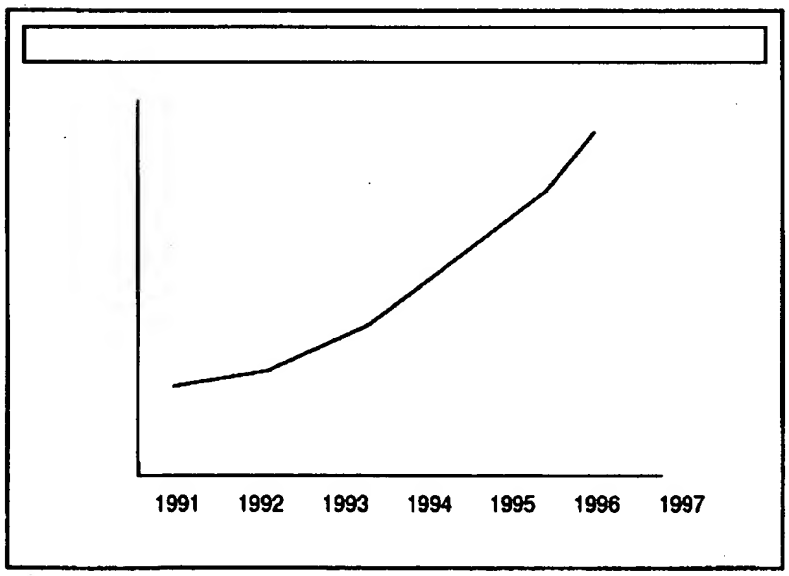


【図２６】



	ドキュメント	リレーション	オブジェクト
検索	特願平10-XXXXXX	特願平10-○○○○○○	
格納		特願平09-○○○○○○	
加工			
流通			

【図２７】



【書類名】　　要約書

【要約】

【課題】　計算量を増大させずに、曖昧パスを含む文書が持つ階層構造に対する多様な検索指定を行うことを可能とした、構造化文書検索方法を提供すること。

【解決手段】　文書の論理構造を含む検索要求に基づいて文書の構造情報を含む検索グラフを生成し（３１，３２）、データの生起位置を特定するデータ生起インデックスと要素名称の生起位置を特定する要素名称生起インデックスを含むインデックスファイル５を利用しながら検索グラフ中において評価可能な部分グラフを優先的に評価する戦略に基づいて検索グラフを巡回することによって最適な検索プランを生成し（３３）、データファイル４に対して検索プランを実行することによって（３４）、検索要求を満足する検索結果を求める。

【選択図】　　図１

出　願　人　履　歴　情　報

識別番号　　〔０００００３０７８〕

１．変更年月日　　１９９０年　８月２２日
［変更理由］　新規登録
住　所　　神奈川県川崎市幸区堀川町７２番地
氏　名　　株式会社東芝